Approximating a Wavefunction as an Unconstrained Sum of Slater Determinants

Gregory Beylkin* Martin J. Mohlenkamp† Fernando Pérez‡

Submitted for publication July 2007.
University of Colorado APPM preprint number 554.

Abstract

The wavefunction for the multiparticle Schrödinger equation is a function of many variables and satisfies an antisymmetry condition, so it is natural to approximate it as a sum of Slater determinants. Many current methods do so, but they impose additional structural constraints on the determinants, such as orthogonality between orbitals or an excitation pattern. We present a method without any such constraints, by which we hope to obtain much more efficient expansions, and insight into the inherent structure of the wavefunction. We use an integral formulation of the problem, a Green's function iteration, and a fitting procedure based on the computational paradigm of separated representations. The core procedure is the construction and solution of a matrix-integral system derived from antisymmetric inner products involving the potential operators. We show how to construct and solve this system with computational complexity competitive with current methods.

AMS Subject Classification: 65Z05 65D15 81-08
Keywords: Multiparticle Schrödinger equation; Slater determinant; curse of dimensionality.

*Department of Applied Mathematics, University of Colorado at Boulder, 526 UCB, Boulder CO 80309-0526; beylkin@colorado.edu.

†Department of Mathematics, Ohio University, 321 Morton Hall, Athens OH 45701; mjm@math.ohiou.edu.

‡Department of Applied Mathematics, University of Colorado at Boulder, 526 UCB, Boulder CO 80309-0526; Fernando.Perez@colorado.edu.

I Introduction

Given the difficulties of solving the multiparticle Schrödinger equation, current numerical methods in quantum chemistry/physics are remarkably successful. Part of their success comes from efficiencies gained by imposing structural constraints on the wavefunction to match physical intuition. However, such methods scale poorly to high accuracy, and are biased to only reveal structures that were part of their own construction. Our goal is to develop a method that scales well to high accuracy and allows an unbiased exploration of the structure of the wavefunction. In this paper we take a step toward this goal by developing a method to approximate the wavefunction as an unconstrained sum of Slater determinants.

Since the multiparticle fermionic wavefunction is an antisymmetric function of many variables, it is natural to approximate it as a sum of Slater determinants, at least as a first step. Motivated by the physical intuition that electrons may be excited into higher energy states, the Configuration Interaction (CI) family of methods choose a set of determinants with predetermined orbitals, and then optimize the coefficients used to combine them. When it is found insufficient, methods to optimize the orbitals, work with multiple reference states, etc., are introduced (along with an alphabet of acronyms). A common feature of all these methods is that they impose some structural constraints on the Slater determinants, such as orthogonality of orbitals or an excitation pattern. As the requested accuracy increases, these structural constraints trigger an explosion in the number of determinants used, making the computation intractable for high accuracy.

The a priori structural constraints present in CI-like methods also force the wavefunction to comply with such structure, whether or not it really is the case. For example, if you use a method that approximates the wavefunction as a linear combination of a reference state and excited states, you could not learn that the wavefunction is better approximated as a linear combination of several non-orthogonal, near-reference states. Thus the choice of numerical method is not just a computational issue; it can help or hinder our understanding of the wavefunction.

For these reasons, our goal is to construct an adaptive numerical method without imposing a priori structural constraints besides that of antisymmetry. In this paper we derive and present an algorithm for approximating a wavefunction with an unconstrained sum of Slater determinants, with fully-adaptive single-electron functions. In particular we discard the notions of reference state and excitation of orbitals. The functions comprising the Slater determinants need not come from a particular basis set, be orthogonal, or follow some excitation pattern. They are computed so as to optimize the overall representation. In this respect we follow the philosophy of separated representations [4, 5], which allow surprisingly accurate expansions with remarkably few terms.

Our construction generates a solution using an iterative procedure based on nonlinear approximations via separated representations. To accomplish this nonlinear approximation, we derive a system of integral equations that describe the fully-correlated many-particle problem. The computational core of the method is the repeated construction and solution of a matrix-integral system of equations.

Specifically, our approach has the following distinctive features:

- We use an adaptive representation for single-electron functions, but our method does not depend on its details.
- We use an integral formulation of the multiparticle Schrödinger equation and a Green's function iteration to converge to the ground-state wavefunction. The Green's function is decomposed and applied using separated approximations obtained by expanding the kernel into Gaussians.

- We use a variant of the so-called alternating least squares algorithm to reduce the error of our approximation using a sum of a given number of Slater determinants.
- We compute antisymmetric inner products involving portions of the Hamiltonian operator by reducing them to formulas involving only combinations of standard integrals. In particular, we avoid the direct application of the electron-electron potential and instead compute convolutions with the Poisson kernel.

By doing this, we hope to represent the effects of correlations in the most natural and concise way possible, thus providing both computational efficiency and physical insight. We believe that this algorithm and the system of integral equations underlying it provide the foundation for a new approach to solving the multiparticle Schrödinger equation. We defer to the sequels several important issues, such as algorithmic size-consistency/extensivity and the treatment of the inter-electron cusp.

In Section II we formulate the problem more carefully, make precise some of the statements that we made in this introduction, and give a high-level description of the method. We then present the derivations and proofs in the following sections.

II Problem Formulation and Description of the Method

II.1 Formulation of the Problem

We consider the time-independent, nonrelativistic, multiparticle Schrödinger equation, and fix the nuclei according to the Born-Oppenheimer approximation, so the equation describes the steady state of an interacting system of electrons. For each of the N electrons in the system there are three spatial variables $\mathbf{r} = (x, y, z)$ and a discrete spin variable σ taking the values $\{-\frac{1}{2}, \frac{1}{2}\}$, which we combine and denote $(\mathbf{r}, \sigma)$ by γ. The Hamiltonian operator $\mathcal{H}$ is a sum of a kinetic energy operator $\mathcal{T}$, a nuclear potential operator $\mathcal{V}$, and an electron-electron interaction operator $\mathcal{W}$, defined in atomic units by

$$\mathcal{H} = \mathcal{T} + \mathcal{V} + \mathcal{W} = -\frac{1}{2}\sum_{i=1}^{N} \Delta_i + \sum_{i=1}^{N} v(\mathbf{r}_i) + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} \frac{1}{\|\mathbf{r}_i - \mathbf{r}_j\|}\,, \tag{1}$$

where Δ_i is the three-dimensional Laplacian acting in the variable $\mathbf{r}_i$ and $v(\mathbf{r})$ is a sum of terms of the form $-Z_a/\|\mathbf{r} - \mathbf{R}_a\|$ from a nucleus at position $\mathbf{R}_a$ with charge Z_a. The antisymmetric eigenfunctions of $\mathcal{H}$ represent electronic states of the system and are called wavefunctions. Antisymmetric means that under the exchange of any two coordinates, the wavefunction is odd, e.g. $\psi(\gamma_2, \gamma_1, \ldots) = -\psi(\gamma_1, \gamma_2, \ldots)$. The bound-state wavefunctions have negative eigenvalues, and are of greatest interest. We will focus on the ground-state wavefunction, which has the most negative eigenvalue, although the techniques can be used for other states. In summary, our goal is to find E and ψ, with E the most negative eigenvalue in

$$\mathcal{H}\psi = E\psi\,, \tag{2}$$

subject to the antisymmetry condition on ψ. Analytic methods can give qualitative results about the solutions, and determine limiting cases, but most quantitative results must be obtained numerically. Although the equation is 'just' an eigenvalue problem, its numerical solution presents several serious difficulties, among them the large number of variables and the antisymmetry condition on the solution. The simplest method that addresses these two difficulties is Hartree-Fock (HF) (see e.g. [28]), which uses the antisymmetrization of a single

product, called a *Slater determinant*, to approximate the N-particle wavefunction, i.e.

$$\psi_{\text{HF}} = \mathcal{A}\prod_{i=1}^{N}\phi_i(\gamma_i) = \frac{1}{N!}\begin{vmatrix} \phi_1(\gamma_1) & \phi_1(\gamma_2) & \cdots & \phi_1(\gamma_N) \\ \phi_2(\gamma_1) & \phi_2(\gamma_2) & \cdots & \phi_2(\gamma_N) \\ \vdots & \vdots & & \vdots \\ \phi_N(\gamma_1) & \phi_N(\gamma_2) & \cdots & \phi_N(\gamma_N) \end{vmatrix}. \tag{3}$$

Any antisymmetric approximation $\tilde{\psi}$ to the wavefunction ψ can be substituted into

$$\frac{\langle \mathcal{H}\tilde{\psi}, \tilde{\psi}\rangle}{\langle \tilde{\psi}, \tilde{\psi}\rangle}, \tag{4}$$

where $\langle\cdot,\cdot\rangle$ is the usual inner product, to obtain an estimate for E. This estimate gives an upper bound on the lowest value of E that solves (2). Substituting (3) into (4), one can iteratively solve for ϕ_i to minimize (4). The resulting ψ_{HF} will best approximate ψ, in the sense of providing the best estimate (4).

To improve upon HF, it is natural to consider the antisymmetrization of a sum of products

$$\psi_{(r)} = \mathcal{A}\sum_{l=1}^{r} s_l \prod_{i=1}^{N}\phi_i^l(\gamma_i), \tag{5}$$

which could also be written as a sum of Slater determinants. The coefficients s_l are introduced in order to have $\|\phi_i^l\| = 1$. Many methods are based on this form, but they use it in different ways. The Configuration Interaction (CI) method (see e.g. [57]) chooses the functions ϕ_i^l from a preselected master set of orthogonal functions and decides on a large number r of combinations to consider, based on excitation level. Substituting (5) into (4) leads to a matrix eigenvalue problem that can be solved for the scalar coefficients s_l. The Multi-Configuration Self-Consistent Field (MCSCF) method (e.g. [20, 11]) solves for the master set of orthogonal functions as well as the scalar coefficients. There are numerous variations and combinations of these methods, too many to describe here.

II.1.1 What is New Here

In this work we construct and demonstrate a method that also uses a wavefunction of the form (5) but without constraints on the ϕ_i^l. We remove both structural constraints, such as an excitation pattern or orthogonality between single-electron functions, and representation constraints, such as those imposed by using a predetermined basis set.

Many methods (e.g. [55, 47, 39, 1, 19, 15, 18, 2, 16, 60, 41]) have loosened the constraints on the Slater determinants in one way or another, often with encouraging results. These works, however, only partially removed the constraints, and so, we claim, did not achieve the full potential of an unconstrained approximation. By removing these constraints we hope to produce much better approximations at much smaller separation rank r than existing methods allow. We also hope to provide new perspective from which to analyze and understand the wavefunction, free from the biases that physical intuition imposes.

Our hopes are based on our work in [4, 5, 43], where we developed general methods to represent and compute with functions and operators in many dimensions. We used sums of separable functions, dubbed *separated representations*, similar to (5). We found rather natural examples where removing constraints produces expansions that are *exponentially* more efficient, i.e. $r = N$ instead of 2^N or $r = \log N$ instead of N. For example, in our approach we can have a two-term representation

$$\psi = \mathcal{A}\prod_{i=1}^{N}\phi_i(\gamma_i) + \mathcal{A}\prod_{i=1}^{N}(\phi_i(\gamma_i) + \phi_{i+N}(\gamma_i)) \tag{6}$$

where $\{\phi_j\}_{j=1}^{2N}$ form an orthogonal set. To represent the same function as (6) while imposing the constraint that factors come from a master orthogonal set would force one to multiply out the second term, and thus use a representation with 2^N terms.

At present we have no proof that the wavefunction is well-approximated by a structure that would benefit from the removal of constraints. The size r needed in practice, and how it depends on the various parameters in the problem, is thus still an open question. In [4, 5, 43], the most interesting examples came from "reverse-engineering" the numerical results to obtain formulas and proofs. We therefore expect that the tools we provide here will allow an exploration of the wavefunction, perhaps revealing unexpected structure, and a strategy for a proof.

II.2 Description of the Algorithm

The removal of constraints in (5), and, thus, the basis sets, coefficients, and other structure that went along with them, also eliminates the conventional strategies for constructing (5) to minimize (4). It leads one to consider how one would compute the ground-state wavefunction if its numerical representation were not an issue. We choose to use an integral iteration, which we sketch in Section II.2.1. In Appendix A we sketch an alternative iteration based on gradient descent.

To use the form (5) we must choose some value of r, which determines the quality of the approximation. In Section II.2.2 we show how to incorporate a nonlinear fitting step within the integral iteration in order to maintain fixed r. Accomplishing this fitting requires a significant amount of machinery, which makes up the body of the paper. Eventually one would want to adaptively determine r, but we do not address that issue here.

II.2.1 A Green's Function Iteration

The eigenvalue equation (2) contains the differential operator $\mathcal{H}$, which has both the discrete negative eigenvalue(s) that we are interested in and unbounded, continuous, positive spectrum. In [31, 32] this differential equation was reformulated as an integral equation, producing an operator with only discrete, bounded spectrum. Such integral formulations are in general far superior to differential formulations, since, e.g. numerical noise is suppressed rather than amplified. An iteration based on the integral formulation with Green's functions was introduced in [31, 32] and used in e.g. [12, 26]. A rigorous analysis of this iteration is given in [44] based on classical theorems from [30, 33, 52, 53, 54]. In this section we review this iteration, and then modify it in Section II.2.2 to preserve our wavefunction representation (5).

Define the Green's function

$$\mathcal{G}_\mu = (\mathcal{T} - \mu\mathcal{I})^{-1}\,, \tag{7}$$

for $\mu < 0$, and consider the Lippmann-Schwinger integral equation

$$\lambda_\mu \psi_\mu = -\mathcal{G}_\mu[(\mathcal{V} + \mathcal{W})\psi_\mu]\,. \tag{8}$$

The subscript μ on λ_μ and ψ_μ are to emphasize the dependence of the eigenvalues and eigenfunctions on μ. The operator $\mathcal{G}_\mu[(\mathcal{V}+\mathcal{W})]$ is compact, so (8) has only discrete spectrum. If $\mu = E$, then there is an eigenvalue $\lambda_\mu = 1$ and the corresponding eigenfunction ψ_μ of (8) is the desired ground-state eigenfunction of (2), as one can see by rearranging (8) into (2). We note that other eigenfunctions may be obtained by deflation.

When $\mu = E$, $\lambda_\mu = 1$ is the largest eigenvalue, so a simple iteration like the power method yields the desired ground-state eigenfunction. The eigenvalues λ_μ depend analytically on μ, so when μ is sufficiently close to E the power method will still yield an eigenfunction of (8)

with energy near the minimum of (4). From ψ_μ and λ_μ one can construct an update rule for μ, based for example on applying Newton's method to solve $\lambda_\mu = 1$.

The convergence rate of the power method to produce ψ_μ and λ_μ is linear, and depends, as usual, on the gap between the eigenvalues in (8). The convergence rate of Newton's method to solve $\lambda_\mu = 1$ is quadratic, so μ will converge to E quadratically, provided that λ_μ and ψ_μ have been found at each step. In the practical use of this approach, one does not wait for the power method to converge at each step, but instead intertwines it with the update of μ. Beginning with an approximation to the energy $\mu_0 \approx E$ and an approximate wavefunction ψ_0, one converts (8) to an iteration

$$\tilde{\psi}_n = -\mathcal{G}_{\mu_n}[(\mathcal{V} + \mathcal{W})\psi_n]\,. \tag{9}$$

After each iteration one normalizes by setting

$$\psi_{n+1} = \tilde{\psi}_n / \|\tilde{\psi}_n\|\,. \tag{10}$$

Following the approach of [26], we can use the update rule

$$\mu_{n+1} = \mu_n - \langle (\mathcal{V} + \mathcal{W})\psi_n, \psi_n - \tilde{\psi}_n \rangle / \|\tilde{\psi}_n\|^2\,, \tag{11}$$

which is equivalent to using Newton's method.

II.2.2 Approximating with Fixed Separation Rank r

We restrict the method to approximate wavefunctions of the form (5), with r fixed, by replacing the definition of $\tilde{\psi}_n$ in (9). We define $\tilde{\psi}_n$ to be the function of the form (5) that minimizes the (least-squares) error

$$\|\tilde{\psi}_n - (-\mathcal{G}_{\mu_n}[(\mathcal{V} + \mathcal{W})\psi_n])\|. \tag{12}$$

Since using (12) instead of (9) introduces an error, the update rule (11) may no longer give quadratic convergence, and in any case is not expected to converge to the true energy. One may choose to replace the update rule (11) with the more robust but slower converging rule

$$\mu_{n+1} = \frac{\langle \mathcal{H}\psi_{n+1}, \psi_{n+1} \rangle}{\|\psi_{n+1}\|^2}\,, \tag{13}$$

which is based on (4). Other rules may be possible as well. At present we do not have enough numerical experience to decide which rule to prefer.

The Green's function iteration itself does not enforce the antisymmetry condition. In order to assure convergence to an antisymmetric solution, we use the pseudo-norm induced by the pseudo inner product $\langle \cdot, \cdot \rangle_{\mathcal{A}} = \langle \mathcal{A}(\cdot), \mathcal{A}(\cdot) \rangle$, as we did in [5].

The least-squares problem (12) is non-linear, and so very difficult in general. To simplify notation in the description of our method, we now suppress the index n in (12) and consider a single problem of that form. We begin by setting $\tilde{\psi} = \psi$, and then iteratively improve $\tilde{\psi}$ to reduce (12). Although we can see several strategies for improving $\tilde{\psi}$, for concreteness we will restrict our description to the strategy most similar to [5]. To improve the approximation $\tilde{\psi}$ we loop through the variables (electrons). The functions in variables other than the current variable are fixed, and the functions in the current variable are modified to minimize the overall error (12). The error (12) depends linearly on the functions in a single variable, so the minimization becomes much easier. This general Alternating Least-Squares (ALS) approach is well-known (see e.g. [27, 36, 38, 10, 14, 58]). Although to minimize (12) one may need to loop through the variables multiple times, it appears to be more cost effective to loop only once and then do the next Green's function iteration. We alternate through the

directions, but for ease of exposition we describe the $k = 1$ case. So, $\tilde{\phi}_k^l$ is fixed for $k > 1$, and we will solve for the values of $\tilde{\phi}_1^l$ for all l.

To refine in the current variable, we set up and solve a linear least-squares problem. The normal equations for a least-squares problem are derived by taking a gradient with respect to the free parameters and setting the result equal to zero. As long as the approximating function is linear and not degenerate in these parameters, the resulting equations are linear and have a unique solution, which minimizes the error with respect to these parameters. Usually these free parameters are coefficients of the representation in some fixed basis. For example, to find the coefficients $\{c_i\}$ to minimize

$$\left\| f - \sum_i c_i g_i \right\|^2 = \left\langle f - \sum_i c_i g_i, f - \sum_i c_i g_i \right\rangle , \tag{14}$$

construct the normal equations

$$\mathbb{A}\mathbf{x} = \mathbf{b} , \tag{15}$$

with

$$A(k,i) = \langle g_k, g_i \rangle \quad \text{and} \quad b(k) = \langle g_k, f \rangle , \tag{16}$$

solve them, and set $c_i = x(i)$. Instead of using coefficients in some basis as our parameters, we take the parameters to be the point values of our functions $\tilde{\phi}_1^l$, so that the gradient becomes a variational derivative. Formally, we consider a basis of delta functions $\delta(\gamma - \cdot)$ and let their coefficients be our parameters. We still obtain linear normal equations (15), but now $\mathbf{b}$ and $\mathbf{x}$ are vectors of functions, and $\mathbb{A}$ is a matrix of integral operators. Specifically, $b(l)$ is a function of γ, $x(l')$ is a function of γ', and $A(l,l')$ is an integral operator mapping functions of γ' to functions of γ. The kernels in $\mathbb{A}$ are formally defined by

$$A(l,l')(\gamma,\gamma') = \tilde{s}_l \tilde{s}_{l'} \left\langle \delta(\gamma - \gamma_1) \prod_{i=2}^{N} \tilde{\phi}_i^l(\gamma_i), \delta(\gamma' - \gamma_1) \prod_{i=2}^{N} \tilde{\phi}_i^{l'}(\gamma_i) \right\rangle_{\mathcal{A}} , \tag{17}$$

and the functions in $\mathbf{b}$ are defined by

$$b(l)(\gamma) = \tilde{s}_l \sum_m^r s_m \left\langle \delta(\gamma - \gamma_1) \prod_{i=2}^{N} \tilde{\phi}_i^l(\gamma_i), -\mathcal{G}_\mu[\mathcal{V} + \mathcal{W}] \prod_{i=1}^{N} \phi_i^m(\gamma_i) \right\rangle_{\mathcal{A}} . \tag{18}$$

Once we solve (15), we set $\tilde{\phi}_1^l = x(l)$. To enforce the normalization convention $\|\tilde{\phi}_1^l\| = 1$ we can divide $\tilde{\phi}_1^l$ by its norm and incorporate the norm into $\tilde{s}_l$.

To solve the matrix-integral system (15), we need an iterative method for solving linear systems that uses only operations compatible with integral operators, such as matrix-vector products, vector scales and additions, and vector inner products. Typically the matrix $\mathbb{A}$ in normal equations is positive-definite. Our operator $\mathbb{A}$ is only semidefinite due to the nullspace in the antisymmetric pseudonorm. Fortunately, $\mathbf{b}$ was computed with the same pseudonorm and has no component in the nullspace of $\mathbb{A}$, so we can still use methods for positive-definite matrices. Based on these considerations, we choose to use the Conjugate Gradient iterative method (see e.g. [21]) to solve (15). One initializes with $\mathbf{r} = \mathbf{b} - \mathbb{A}\mathbf{x}$, $\mathbf{v} = \mathbf{r}$, and $c = \langle \mathbf{r}, \mathbf{r} \rangle$, and then the core of the method is the sequence of assignments $\mathbf{z} \leftarrow \mathbb{A}\mathbf{v}$, $t \leftarrow c/\langle \mathbf{v}, \mathbf{z} \rangle$, $\mathbf{x} \leftarrow \mathbf{x} + t\mathbf{v}$, $\mathbf{r} \leftarrow \mathbf{r} - t\mathbf{z}$, $d \leftarrow \langle \mathbf{r}, \mathbf{r} \rangle$, $\mathbf{v} \leftarrow \mathbf{r} + (d/c)\mathbf{v}$, and $c \leftarrow d$, applied iteratively.

One advantage of using this iterative method with integral operators is that our algorithm does not rely on any particular basis. The representation of $\mathbf{x}$ can naturally be adaptive in γ, for example refining near the nuclei as indicated by the refinement in $\mathbf{b}$. We assume the

availability of some adaptive, high-accuracy representation for single-electron functions, such as the polynomial multiwavelet representation demonstrated in [25, 26], which effectively eliminates the basis-set error. For the estimates of computational cost, we use M to denote the cost to represent a function of γ, or integrate such a function. The antisymmetry constraint requires $N \leq M$, and in general we expect M to be much larger than N.

II.2.3 Summary of the Remainder of the Paper

The core of the paper is the development of the methods needed to construct $\mathbb{A}$ in (17) and $\mathbf{b}$ in (18). First, in Section III, we develop the machinery and algorithms for computing antisymmetric inner products involving the operators $\mathcal{T}$, $\mathcal{V}$, and $\mathcal{W}$. Our formulation uses low-rank perturbations of matrices, thus avoiding cofactor expansions. We also avoid explicit construction of $\mathcal{W}$ by incorporating its effect via spatial convolutions with the Poisson kernel in three dimensions. Second, in Section IV, we show how to compute antisymmetric inner products involving these operators and the delta function $\delta(\gamma - \gamma_1)$. Again the key is to use low-rank perturbations of matrices.

In Section V we assemble all our tools to demonstrate how to perform our main algorithm, and in particular how to construct $\mathbb{A}$ in (17) and $\mathbf{b}$ in (18). We also gather the computational cost for the whole method. The cost depends on the number of electrons N, the separation rank r, the one-particle representation cost M, the number of Green's function iterations I (see Section II.2.1), and the number of conjugate gradient iterations S (see Section II.2.2). Although S in theory could be as many as the number of degrees of freedom rM, we have a very good starting point, and so expect only a very small constant number to be needed. We use $M \log M$ to denote the generic cost to convolve a function of γ with the Poisson kernel $1/\|\mathbf{r}\|$. A Fourier-based Poisson solver on a uniform grid would achieve this complexity; for adaptive methods such as we use it is very difficult to state the cost (see [7, 17]). We use L to denote the number of terms used to approximate the Green's function to relative error ϵ with Gaussians, and prove in Section V.1 that $L = \mathcal{O}((\ln \epsilon)^2)$ independent of μ and N. The final computational cost is then

$$\mathcal{O}(Ir^2N^2[L(N + M \log M) + S(N + M)]). \tag{19}$$

For comparison, the cost to evaluate a single antisymmetric inner product via Löwdin's rules is $\mathcal{O}(N^2(N + M))$.

II.3 Further Considerations

We have implemented the method developed here and tested it sufficiently to verify the correctness of the algorithm as presented. The numerical results are too preliminary to allow us to make any particular claims at this point, however, so we will present them separately. The linear algebra accelerations based on Appendix B have not yet been implemented.

We develop the method in terms of the total variable γ without specifying the spin states. If a specific spin state is imposed on our initial trial wavefunction ψ_0, the iteration will preserve this state.

The representation (5) does not account for the inter-electron cusp (see e.g. [56, 46, 35, 49, 50, 34, 37]), and thus we cannot hope to achieve small error ϵ in the wavefunction with small r. As with Configuration Interaction methods, we may still be able to achieve small error in the energy difference of two systems, which is often the quantity of interest in Chemistry. For the current work, we fix r and adapt $\phi_i^l(\gamma_i)$ and s_l to minimize the error ϵ, rather than fixing ϵ and adaptively determining r. We are developing an extension to (5) that incorporates the cusp, and hope to achieve small error ϵ through it.

Similarly, (5) is not size-consistent/extensive, and thus is not suitable for large systems. We are also developing an extension to (5) suitable for large systems, and hope to achieve linear scaling through it.

Although we have focused on the multiparticle Schrödinger equation, the tools that we have developed are another step towards general-purpose, automatically adaptive methods for solving high-dimensional problems.

III Antisymmetric Inner Products

In this section we develop methods for computing antisymmetric inner products involving $\mathcal{W}$, $\mathcal{V}$, and $\mathcal{T}$. For this purpose, after setting notation, we develop methods for computing with low rank perturbations of matrices, review the antisymmetry constraint and define a notion of maximum coincidence. With these tools we then derive the main formulas.

III.1 Notation

We denote a column *vector* with suppressed indices by $\mathbf{F}$ and with explicit indices by $F(i)$. We denote its conjugate transpose by $\mathbf{F}^*$. We use $\mathbf{e}_i$ to denote the column vector that is one in coordinate i and zero otherwise. A linear *operator* is written $\mathcal{L}$. We denote a *matrix* with suppressed indices by $\mathbb{L}$ and with explicit indices by $L(i,j)$. Recalling that $\mathbf{r} = (x,y,z) \in \mathbf{R}^3$, we combine spatial integration with summation over spins and define the integral

$$\int f(\gamma) d\gamma = \sum_{\sigma \in \{-1/2, 1/2\}} \int f(\mathbf{r}, \sigma) d\mathbf{r}. \tag{20}$$

We define the action of the single-electron kinetic and nuclear potential operators by

$$(\mathcal{T}_* + \mathcal{V}_*)[f](\gamma) = \left(-\frac{1}{2}\Delta + v(\mathbf{r})\right) f(\gamma) = \left(-\frac{1}{2}\Delta + v(\mathbf{r})\right) f(\mathbf{r}, \sigma). \tag{21}$$

In what follows we will reduce the action of the inter-electron potential operator $\mathcal{W}$ to convolutions with the Poisson kernel, so we define

$$\mathcal{W}_{\mathcal{P}}[f](\mathbf{r}) = \int \frac{1}{\|\mathbf{r} - \mathbf{r}'\|} f(\gamma') d\gamma' = \sum_{\sigma' \in \{-1/2, 1/2\}} \int \frac{1}{\|\mathbf{r} - \mathbf{r}'\|} f(\mathbf{r}', \sigma') d\mathbf{r}' \,. \tag{22}$$

We allow these operators to be applied componentwise to vectors and matrices of functions.

Next, we define $\Phi = \prod_{i=1}^N \phi_i(\gamma_i)$, so for example we can write $\langle \tilde{\Phi}, \Phi \rangle_{\mathcal{A}}$ instead of $\left\langle \prod_{i=1}^N \tilde{\phi}_i(\gamma_i), \prod_{i=1}^N \phi_i(\gamma_i) \right\rangle_{\mathcal{A}}$. We also associate with the product Φ a vector of N functions of a single variable,

$$\mathbf{\Phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix}. \tag{23}$$

We can then, for example, construct a new vector of functions $\mathbf{\Theta}$ by applying a matrix to an old one, as in $\mathbf{\Theta} = \mathbb{L}^{-1}\tilde{\mathbf{\Phi}}$. Although we do linear algebra operations on these vectors, we note that $\mathbf{\Phi} + \tilde{\mathbf{\Phi}}$ does not correspond to $\Phi + \tilde{\Phi}$, so there is not a true vector-space structure. Our formulas contain fairly complicated expressions with such vectors, such as $\int \mathbf{\Phi}^* \mathcal{W}_{\mathcal{P}}[\mathbf{\Theta}\mathbf{\Phi}^*]\mathbf{\Theta} d\gamma$. To parse this expression, we note that $\mathbf{\Theta}$ is a column vector of functions and $\mathbf{\Phi}^*$ is a row vector of functions, so $\mathbf{\Theta}\mathbf{\Phi}^*$ is a matrix of functions. Then $\mathcal{W}_{\mathcal{P}}[\mathbf{\Theta}\mathbf{\Phi}^*]$ is

still a matrix of functions, but applying $\mathbf{\Phi}^*$ on the left and $\mathbf{\Theta}$ on the right yields a single function, which is integrated in the implied variable γ to yield a number. When explicit specification of the variable involved is needed, the notation $\mathbf{\Phi}(\gamma)$ indicates that the single variable γ is used in all the functions.

III.2 Determinants of Low-Rank Perturbations of Matrices

Since the antisymmetric inner product involves determinants, we will use some linear algebra relations for them. Proposition 2 in this section is used heavily, and is the key to avoiding rather unpleasant cofactor expansions.

Proposition 1 (Determinant via Schur Complement) *Let $\mathbb{A}$ be a nonsingular square matrix, $\mathbb{D}$ a square matrix, and $\mathbb{B}$ and $\mathbb{C}$ matrices of appropriate size. Then*

$$\begin{vmatrix} \mathbb{A} & \mathbb{B} \\ \mathbb{C} & \mathbb{D} \end{vmatrix} = |\mathbb{A}| \left|\mathbb{D} - \mathbb{C}\mathbb{A}^{-1}\mathbb{B}\right| . \tag{24}$$

Proof: (see e.g. [51]) It is easy to verify directly that

$$\begin{bmatrix} \mathbb{A} & \mathbb{B} \\ \mathbb{C} & \mathbb{D} \end{bmatrix} = \begin{bmatrix} \mathbb{I} & 0 \\ \mathbb{C}\mathbb{A}^{-1} & \mathbb{I} \end{bmatrix} \begin{bmatrix} \mathbb{A} & 0 \\ 0 & \mathbb{D} - \mathbb{C}\mathbb{A}^{-1}\mathbb{B} \end{bmatrix} \begin{bmatrix} \mathbb{I} & \mathbb{A}^{-1}\mathbb{B} \\ 0 & \mathbb{I} \end{bmatrix}. \tag{25}$$

Since the determinants of the first and third matrices are equal to one, the determinant of the middle matrix gives the desired result. □

Proposition 2 (Determinant of a Perturbation of the Identity) *Let $\{\mathbf{u}_q\}_{q=1}^Q$ and $\{\mathbf{v}_q\}_{q=1}^Q$ be two sets of vectors of the same length, and $\mathbf{u}_q\mathbf{v}_q^*$ denote the outer product of $\mathbf{u}_q$, and $\mathbf{v}_q$. Then*

$$\left|\mathbb{I} + \sum_{q=1}^{Q} \mathbf{u}_q\mathbf{v}_q^*\right| = \begin{vmatrix} 1+\mathbf{v}_1^*\mathbf{u}_1 & \mathbf{v}_1^*\mathbf{u}_2 & \cdots & \mathbf{v}_1^*\mathbf{u}_Q \\ \mathbf{v}_2^*\mathbf{u}_1 & 1+\mathbf{v}_2^*\mathbf{u}_2 & \cdots & \mathbf{v}_2^*\mathbf{u}_Q \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_Q^*\mathbf{u}_1 & \mathbf{v}_Q^*\mathbf{u}_2 & \cdots & 1+\mathbf{v}_Q^*\mathbf{u}_Q \end{vmatrix}. \tag{26}$$

Proof: Let $\mathbb{U}$ be the matrix with the vectors $\{\mathbf{u}_q\}$ as its columns, and $\mathbb{V}$ the matrix with the vectors $\{\mathbf{v}_q\}$ as its columns. Note that $\mathbb{U}$ and $\mathbb{V}$ are of the same size. By Proposition 1 we have

$$\begin{vmatrix} \mathbb{I} & \mathbb{U} \\ -\mathbb{V}^* & \mathbb{I} \end{vmatrix} = |\mathbb{I} + \mathbb{V}^*\mathbb{U}| , \tag{27}$$

which evaluates to the right side of (26). Exchanging the roles of $\mathbb{A}$ and $\mathbb{D}$ in Proposition 1 we have

$$\begin{vmatrix} \mathbb{I} & \mathbb{U} \\ -\mathbb{V}^* & \mathbb{I} \end{vmatrix} = |\mathbb{I} + \mathbb{U}\mathbb{V}^*| , \tag{28}$$

which evaluates to the left side of (26). □

The $Q = 1$ case is well-known (see e.g. [51]) but we have not found the general case in the literature.

III.3 The Modified Pseudo-inverse

The singular value decomposition (SVD) (e.g. [21]) of a $N \times N$ matrix is

$$\mathbb{A} = \sum_{i=1}^{N} s_i \mathbf{u}_i \mathbf{v}_i^* = \mathbb{U}\mathbb{S}\mathbb{V}^* \,, \tag{29}$$

where the matrices $\mathbb{U}$ and $\mathbb{V}$ are unitary and the singular values $\{s_i\}$ are non-neganive and in descending order. The left singular vectors $\{\mathbf{u}_i\}$ form an orthonormal set, as do the right singular vectors $\{\mathbf{v}_i\}$. The pseudo-inverse is defined as

$$\mathbb{A}^\dagger = \sum_{i=1}^{N-Q} s_i^{-1} \mathbf{v}_i \mathbf{u}_i^* \,, \tag{30}$$

where Q is the dimension of the (numerical) nullspace. We also define a projection matrix onto the nullspace

Definition 3

$$\mathbb{A}^\perp = \sum_{i=N-Q+1}^{N} \mathbf{v}_i \mathbf{u}_i^* \tag{31}$$

and a modified pseudo-inverse

Definition 4 (Modified Pseudo-Inverse)

$$\mathbb{A}^\ddagger = \mathbb{A}^\dagger + \mathbb{A}^\perp \,. \tag{32}$$

Note that $\mathbb{A}^\perp$ and thus $\mathbb{A}^\ddagger$ are not uniquely defined since the choice of basis for the nullspace is not unique. For our purposes any consistent choice works. The modified pseudo-inverse behaves much like the pseudo-inverse, but always has a non-zero determinant,

$$|\mathbb{A}^\ddagger| = \left(|\mathbb{U}||\mathbb{V}^*| \prod_{s_i \neq 0} s_i \right)^{-1} \neq 0 \,. \tag{33}$$

III.4 The Antisymmetrizer and Löwdin's Rule

Given a separable function, its antisymmetric projection can be found by applying the *antisymmetrizer* $\mathcal{A}$ (see e.g. [48]), also called the *skew-symmetrization* or *alternation* (see e.g. [45, 51]), resulting in a Slater determinant. In the vector notation (23), we have

$$\mathcal{A}\Phi = \frac{1}{N!} \left| \begin{bmatrix} \mathbf{\Phi}(\gamma_1) & \cdots & \mathbf{\Phi}(\gamma_N) \end{bmatrix} \right| = \frac{1}{N!} \begin{vmatrix} \phi_1(\gamma_1) & \phi_1(\gamma_2) & \cdots & \phi_1(\gamma_N) \\ \phi_2(\gamma_1) & \phi_2(\gamma_2) & \cdots & \phi_2(\gamma_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(\gamma_1) & \phi_N(\gamma_2) & \cdots & \phi_N(\gamma_N) \end{vmatrix} . \tag{34}$$

One cannot explicitly form a Slater determinant $\mathcal{A}\Phi$ for large N since it would have $N!$ terms. However, one can compute the antisymmetric pseudo inner product

$$\langle \tilde{\Phi}, \Phi \rangle_{\mathcal{A}} \stackrel{\text{def}}{=} \langle \mathcal{A}\tilde{\Phi}, \mathcal{A}\Phi \rangle = \langle \tilde{\Phi}, \mathcal{A}\Phi \rangle = \langle \mathcal{A}\tilde{\Phi}, \Phi \rangle, \tag{35}$$

where the first equality is a definition and the others follow since $\mathcal{A}$ is an orthogonal projector. It is not a true inner product because it has a nullspace. To compute (35), first construct the matrix $\mathbb{L}$ with entries

$$L(i,j) = \langle \tilde{\phi}_i, \phi_j \rangle \tag{36}$$

at cost $\mathcal{O}(N^2 M)$. Then use $\langle \tilde{\Phi}, \Phi \rangle_{\mathcal{A}} = \langle \mathcal{A}\tilde{\Phi}, \Phi \rangle$ and move the integrals inside the determinant to obtain

$$\langle \tilde{\Phi}, \Phi \rangle_{\mathcal{A}} = \frac{1}{N!} |\mathbb{L}| \,, \tag{37}$$

which is the so-called Löwdin's rule (e.g. [40, 48]). Since $\mathbb{L}$ is an ordinary matrix, its determinant can be computed with cost $\mathcal{O}(N^3)$ (or less). The denominator $N!$ need never be computed, since it will occur in every term in our equations, and so cancels.

Our method for enforcing the antisymmetry constraint, as described in [5], is to use the pseudo-norm based on the antisymmetric inner product $\langle \cdot, \cdot \rangle_{\mathcal{A}}$ for the least-squares fitting (12).

III.5 Maximum Coincidence

Consider two products, $\Phi = \prod_{i=1}^{N} \phi_i(\gamma_i)$ and $\tilde{\Phi} = \prod_{i=1}^{N} \tilde{\phi}_i(\gamma_i)$, stored in the vector notation of (23) as $\mathbf{\Phi}$ and $\tilde{\mathbf{\Phi}}$. To specify which functions were used to compute $\mathbb{L}$ in (36), we use the notation $\mathbb{L}(\tilde{\Phi}, \Phi)$. The matrix of inner products $\mathbb{L} = \mathbb{L}(\tilde{\Phi}, \Phi)$ is in general full. Defining

$$\mathbf{\Theta} = \mathbb{L}^{-1} \tilde{\mathbf{\Phi}} \,, \tag{38}$$

we have

$$\begin{aligned} \mathcal{A}\Theta = \frac{1}{N!} \left| \begin{bmatrix} (\mathbb{L}^{-1}\tilde{\mathbf{\Phi}})(\gamma_1) & \cdots & (\mathbb{L}^{-1}\tilde{\mathbf{\Phi}})(\gamma_N) \end{bmatrix} \right| \\ = |\mathbb{L}^{-1}| \frac{1}{N!} \left| \begin{bmatrix} \tilde{\mathbf{\Phi}}(\gamma_1) & \cdots & \tilde{\mathbf{\Phi}}(\gamma_N) \end{bmatrix} \right| = |\mathbb{L}^{-1}| \mathcal{A}\tilde{\Phi} \,. \end{aligned} \tag{39}$$

Thus the antisymmetrizations of $\tilde{\Phi}$ and Θ are the same up to a constant, and we can use Θ instead of $\tilde{\Phi}$ in calculations. The advantage of using Θ is that the resulting matrix of inner products $\hat{\mathbb{L}} = \mathbb{L}(\Theta, \Phi) = \mathbb{I}$; in other words, we have the biorthogonality property $\langle \theta_i, \phi_j \rangle = \delta_{ij}$. To show this, write the matrix $\hat{\mathbb{L}}$ as $\int \mathbf{\Theta}\mathbf{\Phi}^* d\gamma$, where the integration is elementwise. Substituting for $\mathbf{\Theta}$, we have $\int (\mathbb{L}^{-1}\tilde{\mathbf{\Phi}})\mathbf{\Phi}^* d\gamma$. Since the integration is elementwise it commutes with $\mathbb{L}^{-1}$ and we have $\mathbb{L}^{-1} \int \tilde{\mathbf{\Phi}}\mathbf{\Phi}^* d\gamma = \mathbb{L}^{-1}\mathbb{L} = \mathbb{I}$. The computational cost to construct $\mathbf{\Theta}$ is $\mathcal{O}(N^2(N+M))$.

When the matrix $\mathbb{L}$ in (36) is singular, we define $\mathbf{\Theta} = \mathbb{L}^{\ddagger}\tilde{\mathbf{\Phi}}$ using the modified pseudo-inverse of Definition 4. By the same argument as before, we have $|\mathbb{L}^{\ddagger}|^{-1}\mathcal{A}\Theta = \mathcal{A}\tilde{\Phi}$. The matrix $\int \mathbf{\Theta}\mathbf{\Phi}^* d\gamma$ evaluates to $\mathbb{L}^{\ddagger}\mathbb{L} = \mathbb{I} - \sum_{i=N-Q+1}^{N} \mathbf{v}_i \mathbf{v}_i^*$. For notational convenience in later sections, we will re-index our singular values and vectors so that the first Q generate the nullspace, rather than the last Q.

Remark 5 *Within Configuration Interaction methods, the functions in $\mathbf{\Phi}$ and $\tilde{\mathbf{\Phi}}$ are taken from a master set of orthonormal functions, and $\mathbf{\Theta}$ is simply a signed permutation of $\tilde{\mathbf{\Phi}}$ so that $\phi_j = \theta_j$ for as many j as possible. This is known as the 'maximum coincidence' ordering. The construction we use generalizes this notion.*

III.6 Antisymmetric Inner Product with the Electron-Electron Potential $\mathcal{W}$ Present

In this section we derive formulas for computing antisymmetric inner products that include the electron-electron interaction potential. Although the derivation is somewhat messy, the resulting formulas are rather clean, and we use them verbatim in the computations. The main ideas are given in this section, and then reused in later sections for other cases.

Proposition 6 *When $\mathbb{L}$ from (36) is nonsingular,*

$$\left\langle \tilde{\Phi}, \mathcal{W}\Phi \right\rangle_{\mathcal{A}} \stackrel{\text{def}}{=} \left\langle \mathcal{A} \prod_{j=1}^{N} \tilde{\phi}_j(\gamma_j), \left(\frac{1}{2} \sum_{i \neq j} \frac{1}{\|\mathbf{r}_i - \mathbf{r}_j\|} \right) \prod_{j=1}^{N} \phi_j(\gamma_j) \right\rangle \tag{40}$$

is equal to

$$\frac{1}{2} \frac{|\mathbb{L}|}{N!} \int \mathbf{\Phi}^* \mathbf{\Theta} \mathcal{W}_{\mathcal{P}} \left[\mathbf{\Phi}^* \mathbf{\Theta} \right] - \mathbf{\Phi}^* \mathcal{W}_{\mathcal{P}} \left[\mathbf{\Theta} \mathbf{\Phi}^* \right] \mathbf{\Theta} d\gamma \,, \tag{41}$$

where $\mathbf{\Theta} = \mathbb{L}^{-1} \tilde{\mathbf{\Phi}}$.

Proof: Using the maximum-coincidence procedure in Section III.5, (40) is equal to $|\mathbb{L}| \langle \Theta, \mathcal{W}\Phi \rangle_{\mathcal{A}}$. We reorganize and find that we must compute

$$\frac{1}{2} \frac{|\mathbb{L}|}{N!} \int \left(\sum_{i \neq j} \frac{1}{\|\mathbf{r}_i - \mathbf{r}_j\|} \right) \prod_{j=1}^{N} \overline{\phi}_j(\gamma_j) \begin{vmatrix} \theta_1(\gamma_1) & \theta_1(\gamma_2) & \cdots & \theta_1(\gamma_N) \\ \theta_2(\gamma_1) & \theta_2(\gamma_2) & \cdots & \theta_2(\gamma_N) \\ \vdots & \vdots & \ddots & \vdots \\ \theta_N(\gamma_1) & \theta_N(\gamma_2) & \cdots & \theta_N(\gamma_N) \end{vmatrix} d\gamma_1 \cdots d\gamma_N \,. \tag{42}$$

By moving the sum outside of the integral, we can integrate in all directions except γ_i and γ_j. Using $\langle \theta_m, \phi_n \rangle = \delta_{mn}$, we obtain

$$\begin{aligned} &\frac{1}{2} \frac{|\mathbb{L}|}{N!} \sum_{i \neq j} \int \frac{1}{\|\mathbf{r} - \mathbf{r}'\|} \begin{vmatrix} 1 & \cdots & \overline{\phi}_i(\gamma)\theta_1(\gamma) & \cdots & \overline{\phi}_j(\gamma')\theta_1(\gamma') & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \overline{\phi}_i(\gamma)\theta_i(\gamma) & \cdots & \overline{\phi}_j(\gamma')\theta_i(\gamma') & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & \overline{\phi}_i(\gamma)\theta_j(\gamma) & \cdots & \overline{\phi}_j(\gamma')\theta_j(\gamma') & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & \overline{\phi}_i(\gamma)\theta_N(\gamma) & \cdots & \overline{\phi}_j(\gamma')\theta_N(\gamma') & \cdots & 1 \end{vmatrix} d\gamma d\gamma' \\ &= \frac{1}{2} \frac{|\mathbb{L}|}{N!} \sum_{i \neq j} \int \frac{1}{\|\mathbf{r} - \mathbf{r}'\|} \left| \mathbb{I} + \left(\overline{\phi}_i(\gamma) \mathbf{\Theta}(\gamma) - \mathbf{e}_i \right) \mathbf{e}_i^* + \left(\overline{\phi}_j(\gamma') \mathbf{\Theta}(\gamma') - \mathbf{e}_j \right) \mathbf{e}_j^* \right| d\gamma d\gamma' \,. \end{aligned} \tag{43}$$

Since the inner matrix is a low-rank perturbation of the identity, we reduce its determinant using Proposition 2 and obtain

$$\frac{1}{2} \frac{|\mathbb{L}|}{N!} \sum_{i \neq j} \int \frac{1}{\|\mathbf{r} - \mathbf{r}'\|} \overline{\phi}_i(\gamma) \overline{\phi}_j(\gamma') \begin{vmatrix} \theta_i(\gamma) & \theta_i(\gamma') \\ \theta_j(\gamma) & \theta_j(\gamma') \end{vmatrix} d\gamma d\gamma' \,. \tag{44}$$

The determinant is zero if $j = i$, so we do not need to explicitly prohibit it as we needed to in (43) and above. The antisymmetrization has caused a convenient cancellation of a

fictitious self-interaction, and, thus, allowed us to decouple the two sums. Expanding out the determinant and rearranging the terms, we obtain

$$\frac{1}{2}\frac{|\mathbb{L}|}{N!}\int\left(\sum_i \overline{\phi}_i(\gamma)\theta_i(\gamma)\right)\left[\int\frac{1}{\|\mathbf{r}-\mathbf{r}'\|}\left(\sum_j \overline{\phi}_j(\gamma')\theta_j(\gamma')\right)d\gamma'\right]d\gamma \\ -\frac{1}{2}\frac{|\mathbb{L}|}{N!}\int\sum_i\sum_j \overline{\phi}_i(\gamma)\theta_j(\gamma)\left[\int\frac{1}{\|\mathbf{r}-\mathbf{r}'\|}\overline{\phi}_j(\gamma')\theta_i(\gamma')d\gamma'\right]d\gamma\,. \quad (45)$$

In our compact notation, this yields (41). □

We now consider the computational cost of (41). In the first term in (41), computing $\mathbf{\Phi}^*\mathbf{\Theta}$ costs $\mathcal{O}(NM)$, applying $\mathcal{W}_{\mathcal{P}}[\cdot]$ to it costs $\mathcal{O}(M\log M)$, and the integral in γ costs $\mathcal{O}(M)$. In the second term, $\mathbf{\Phi}\mathbf{\Theta}^*$ costs $\mathcal{O}(N^2M)$, applying $\mathcal{W}_{\mathcal{P}}[\cdot]$ to it costs $\mathcal{O}(N^2M\log M)$, applying $\mathbf{\Theta}^*$ and then $\mathbf{\Phi}$ costs $\mathcal{O}(N^2M)$, and then the integral in γ costs $\mathcal{O}(M)$. Including the cost to construct $\mathbf{\Theta}$, our total cost is $\mathcal{O}(N^2(N+M\log M))$.

III.6.1 The Singular Case

In this section we investigate the case when the matrix $\mathbb{L}$ from (36) is singular. Inserting the definition $\mathbf{\Theta}=\mathbb{L}^{-1}\tilde{\mathbf{\Phi}}$ into our main formula (41), we have

$$\frac{1}{2}\frac{|\mathbb{L}|}{N!}\int\mathbf{\Phi}^*\mathbb{L}^{-1}\tilde{\mathbf{\Phi}}\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\mathbb{L}^{-1}\tilde{\mathbf{\Phi}}\right]-\mathbf{\Phi}^*\mathcal{W}_{\mathcal{P}}\left[\mathbb{L}^{-1}\tilde{\mathbf{\Phi}}\mathbf{\Phi}^*\right]\mathbb{L}^{-1}\tilde{\mathbf{\Phi}}d\gamma\,. \quad (46)$$

In terms of the SVD (29), we can express

$$\mathbb{L}^{-1}=\sum_{j=1}^{N}s_j^{-1}\mathbf{v}_j\mathbf{u}_j^* \quad\text{and}\quad |\mathbb{L}|=|\mathbb{U}||\mathbb{V}^*|\prod_i s_i\,. \quad (47)$$

Inserting these expressions into (46), we have

$$\frac{1}{2}\frac{|\mathbb{U}||\mathbb{V}^*|\prod_i s_i}{N!}\int\mathbf{\Phi}^*\sum_{j=1}^{N}s_j^{-1}\mathbf{v}_j\mathbf{u}_j^*\tilde{\mathbf{\Phi}}\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\sum_{k=1}^{N}s_k^{-1}\mathbf{v}_k\mathbf{u}_k^*\tilde{\mathbf{\Phi}}\right] \\ -\mathbf{\Phi}^*\mathcal{W}_{\mathcal{P}}\left[\sum_{j=1}^{N}s_j^{-1}\mathbf{v}_j\mathbf{u}_j^*\tilde{\mathbf{\Phi}}\mathbf{\Phi}^*\right]\sum_{k=1}^{N}s_k^{-1}\mathbf{v}_k\mathbf{u}_k^*\tilde{\mathbf{\Phi}}d\gamma \\ =\frac{1}{2}\frac{|\mathbb{U}||\mathbb{V}^*|}{N!}\sum_{j=1}^{N}\sum_{k=1}^{N}\prod_{i\neq j,k}s_i\int\mathbf{\Phi}^*\mathbf{v}_j\mathbf{u}_j^*\tilde{\mathbf{\Phi}}\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\mathbf{v}_k\mathbf{u}_k^*\tilde{\mathbf{\Phi}}\right] \\ -\mathbf{\Phi}^*\mathbf{v}_j\mathcal{W}_{\mathcal{P}}\left[\mathbf{u}_j^*\tilde{\mathbf{\Phi}}\mathbf{\Phi}^*\mathbf{v}_k\right]\mathbf{u}_k^*\tilde{\mathbf{\Phi}}d\gamma\,. \quad (48)$$

If $\mathbb{L}$ is singular then at least one s_i is zero, and only terms that exclude those from the product in (48) are nonzero. Since we exclude two indices in the product, if more than two s_i are zero then the entire inner product is zero. If exactly two are zero then only one term in the sum survives. If exactly one is zero then we can simplify from a double to a single sum, using symmetry. Recalling the modified pseudo inverse from Definition 4 and sorting the zero s_i to the beginning for notational convenience, we obtain the following propositions.

Proposition 7 *When the rank-deficiency of* $\mathbb{L}$ *is more than two, the antisymmetric inner product (40) evaluates to zero.*

Proposition 8 *When the rank-deficiency of* $\mathbb{L}$ *is equal to two, the antisymmetric inner product (40) is equal to*

$$\frac{1}{|\mathbb{L}^{\ddagger}|N!}\int \mathbf{\Phi}^*\mathbf{v}_1\mathbf{u}_1^*\tilde{\mathbf{\Phi}}\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\mathbf{v}_2\mathbf{u}_2^*\tilde{\mathbf{\Phi}}\right] - \mathbf{\Phi}^*\mathbf{v}_1\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\mathbf{v}_2\mathbf{u}_1^*\tilde{\mathbf{\Phi}}\right]\mathbf{u}_2^*\tilde{\mathbf{\Phi}}d\gamma\,. \tag{49}$$

Proposition 9 *When the rank-deficiency of* $\mathbb{L}$ *is equal to one, defining* $\mathbf{\Theta} = \mathbb{L}^{\dagger}\tilde{\mathbf{\Phi}}$ *or* $\mathbf{\Theta} = \mathbb{L}^{\ddagger}\tilde{\mathbf{\Phi}}$*, the antisymmetric inner product (40) is equal to*

$$\frac{1}{|\mathbb{L}^{\ddagger}|N!}\int \mathbf{\Phi}^*\mathbf{v}_1\mathbf{u}_1^*\tilde{\mathbf{\Phi}}\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\mathbf{\Theta}\right] - \mathbf{\Phi}^*\mathbf{v}_1\mathcal{W}_{\mathcal{P}}\left[\mathbf{u}_1^*\tilde{\mathbf{\Phi}}\mathbf{\Phi}^*\right]\mathbf{\Theta}d\gamma\,. \tag{50}$$

In computing (49), constructing $\mathbf{\Phi}^*\mathbf{v}_1$, $\mathbf{\Phi}^*\mathbf{v}_2$, $\mathbf{u}_1^*\tilde{\mathbf{\Phi}}$, and $\mathbf{u}_2^*\tilde{\mathbf{\Phi}}$ costs $\mathcal{O}(NM)$, applying $\mathcal{W}_{\mathcal{P}}[\cdot]$ costs $\mathcal{O}(M\log M)$ and, finally, the integral in γ costs $\mathcal{O}(M)$. In computing (50), the first term costs $\mathcal{O}(NM)$ to form $\mathbf{\Phi}^*\mathbf{\Theta}$, $\mathcal{O}(M\log M)$ to apply $\mathcal{W}_{\mathcal{P}}[\cdot]$, and $\mathcal{O}(M)$ to integrate in γ. The second term costs $\mathcal{O}(NM)$ to form $\mathbf{u}_1^*\tilde{\mathbf{\Phi}}\mathbf{\Phi}$, $\mathcal{O}(NM\log M)$ to apply $\mathcal{W}_{\mathcal{P}}[\cdot]$, $\mathcal{O}(NM)$ to apply $\mathbf{\Theta}$, and $\mathcal{O}(M)$ to integrate in γ. In total, the computational cost for the singular cases are less than the cost of the nonsingular case.

Remark 10 *In the Configuration Interaction context, rank-deficiency two corresponds to a double excitation. The vectors* $\mathbf{u}_i$ *and* $\mathbf{v}_i$ *would be zero except for a single entry, and so select the locations of the excited electrons out of* $\mathbf{\Phi}$ *and* $\tilde{\mathbf{\Phi}}$*. Proposition 8 then reduces to the Slater-Condon rules [13].*

III.7 Antisymmetric Inner Product with $\mathcal{T}$ and/or $\mathcal{V}$ Present

Since $\mathcal{T}$ and $\mathcal{V}$ both have the structure of a sum of one-directional operators, we state the formulas for their sum, although of course they can be treated individually.

Proposition 11 *If* $\mathbb{L}$ *from (36) is nonsingular,*

$$\left\langle\tilde{\Phi},(\mathcal{T}+\mathcal{V})\Phi\right\rangle_{\mathcal{A}} \stackrel{\text{def}}{=} \left\langle\mathcal{A}\prod_{j=1}^{N}\tilde{\phi}_j(\gamma_j),\left(\sum_i -\frac{1}{2}\Delta_i + v(\mathbf{r}_i)\right)\prod_{j=1}^{N}\phi_j(\gamma_j)\right\rangle \tag{51}$$

is equal to

$$\frac{|\mathbb{L}|}{N!}\int (\mathcal{T}_* + \mathcal{V}_*)\left[\mathbf{\Phi}\right]^*\mathbf{\Theta}d\gamma\,. \tag{52}$$

Proof: We follow the same procedure as we used for the electron-electron operator $\mathcal{W}$ in Section III.6. Instead of (43) we have the simpler expression

$$\frac{|\mathbb{L}|}{N!}\sum_i\int\left|\mathbb{I} + \left((\mathcal{T}_* + \mathcal{V}_*)\left[\overline{\phi}_i\right](\gamma)\mathbf{\Theta}(\gamma) - \mathbf{e}_i\right)\mathbf{e}_i^*\right|d\gamma\,. \tag{53}$$

Applying Proposition 2 we obtain (52). □

To analyze the computational cost to compute (52), we note that it costs $\mathcal{O}(NM)$ to apply $(\mathcal{T}_* + \mathcal{V}_*)[\cdot]$. Including the cost for the maximum coincidence transformation, our total cost is thus $\mathcal{O}(N^2(N+M))$.

III.7.1 The Singular Case

We now state the formula when $\mathbb{L}$ is singular. The analysis is similar to that for $\mathcal{W}$ in Section III.6.1.

Proposition 12 *If the rank-deficiency of* $\mathbb{L}$ *is greater than one, (51) evaluates to zero. If it is equal to one we have*

$$\frac{1}{|\mathbb{L}^\ddagger| N!} \int (\mathcal{T}_* + \mathcal{V}_*)\,[\mathbf{\Phi}^* \mathbf{v}_1] \mathbf{u}_1^* \tilde{\mathbf{\Phi}} d\gamma\,. \tag{54}$$

To compute (54), it costs $\mathcal{O}(NM)$ to form $\mathbf{\Phi}^*\mathbf{v}_1$ and $\mathbf{u}_1^*\tilde{\mathbf{\Phi}}$, and $\mathcal{O}(M)$ to apply $(\mathcal{T}_* + \mathcal{V}_*)\,[\cdot]$.

IV Incorporating Delta Functions into the Antisymmetric Inner Products

In this section we show how to compute antisymmetric inner products when one of the component functions is replaced by a delta function. For concreteness, we will replace $\tilde{\phi}_1(\gamma_1)$ by $\delta(\gamma - \gamma_1)$.

IV.1 Löwdin's Rule with $\delta(\gamma - \gamma_1)$ Present

The matrix $\mathbb{L}$ from (36) is defined by $L(i,j) = \langle \tilde{\phi}_i, \phi_j \rangle$. If we replace $\tilde{\phi}_1(\gamma_1)$ by $\delta(\gamma - \gamma_1)$, then the first row depends on γ and is given by $L(1,j) = \langle \delta(\gamma - \cdot), \phi_j \rangle = \phi_j(\gamma)$. We thus have a matrix that depends on γ,

$$\mathbb{L}(\gamma) = \begin{bmatrix} \phi_1(\gamma) & \phi_2(\gamma) & \cdots & \phi_N(\gamma) \\ \langle \tilde{\phi}_2, \phi_1 \rangle & \langle \tilde{\phi}_2, \phi_2 \rangle & \cdots & \langle \tilde{\phi}_2, \phi_N \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \tilde{\phi}_N, \phi_1 \rangle & \langle \tilde{\phi}_N, \phi_2 \rangle & \cdots & \langle \tilde{\phi}_N, \phi_N \rangle \end{bmatrix}. \tag{55}$$

To compute with $\mathbb{L}(\gamma)$ without resorting to cofactor expansions, we express $\mathbb{L}(\gamma)$ as a rank-one perturbation of a matrix of numbers. Define

$$\mathbb{E} = \begin{bmatrix} \overline{d}(1) & \overline{d}(2) & \cdots & \overline{d}(N) \\ \langle \tilde{\phi}_2, \phi_1 \rangle & \langle \tilde{\phi}_2, \phi_2 \rangle & \cdots & \langle \tilde{\phi}_2, \phi_N \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \tilde{\phi}_N, \phi_1 \rangle & \langle \tilde{\phi}_N, \phi_2 \rangle & \cdots & \langle \tilde{\phi}_N, \phi_N \rangle \end{bmatrix}, \tag{56}$$

where the vector $\mathbf{d}^*$ is chosen to be a unit vector orthogonal to the remaining rows of $\mathbb{E}$. This choice assures that the rank deficiency of $\mathbb{E}$ will be smaller than or equal to the rank deficiency of the matrix with any other first row. It also gives us some convenient properties, namely $\mathbb{E}\mathbf{d} = \mathbf{e}_1$, $\mathbf{d}^*\mathbb{E}^\ddagger = \mathbf{e}_1^*$, $\mathbb{E}^\ddagger \mathbf{e}_1 = \mathbf{d}$, and $\mathbf{e}_1^*\mathbb{E} = \mathbf{d}^*$, where $\mathbb{E}^\ddagger$ is the modified pseudo-inverse of Definition 4. It costs $\mathcal{O}(N^2M)$ to construct $\mathbb{E}$ and $\mathcal{O}(N^3)$ to compute $\mathbb{E}^\ddagger$ and $|\mathbb{E}|$.

We then have

$$\mathbb{L}(\gamma) = \mathbb{E} + \mathbf{e}_1(\mathbf{\Phi}(\gamma) - \mathbf{d})^* \tag{57}$$

and, with the help of Proposition 2, compute

$$|\mathbb{L}(\gamma)| = |\mathbb{E}||\mathbb{I} + \mathbf{d}(\mathbf{\Phi}(\gamma) - \mathbf{d})^*| = |\mathbb{E}|\,(1 + (\mathbf{\Phi}(\gamma) - \mathbf{d})^*\mathbf{d}) = |\mathbb{E}|\,\mathbf{\Phi}(\gamma)^*\mathbf{d}\,, \tag{58}$$

which yields

Proposition 13

$$\left\langle \delta(\gamma-\gamma_1)\prod_{i=2}^{N}\tilde{\phi}_i(\gamma_i), \prod_{i=1}^{N}\phi_i(\gamma_i)\right\rangle_{\mathcal{A}} = |\mathbb{E}|\,\mathbf{\Phi}(\gamma)^*\mathbf{d}\,, \tag{59}$$

where $\mathbb{E}$ *and* $\mathbf{d}$ *are defined as above.*

Remark 14 *If* $i>1$ *then*

$$\langle |\mathbb{E}|\,\mathbf{\Phi}^*\mathbf{d}, \tilde{\phi}_i\rangle = |\mathbb{E}|\langle \mathbf{\Phi}, \tilde{\phi}_i\rangle^*\mathbf{d} = |\mathbb{E}|E(i,\cdot)^*\mathbf{d} = 0\,, \tag{60}$$

since $\mathbf{d}$ *is orthogonal to* $E(i,\cdot)$*, which is row number* i *of* $\mathbb{E}$*. Thus the function (59) is orthogonal to* $\tilde{\phi}_i$ *for* $i>1$*. The same property will hold when the operators* $\mathcal{T}$*,* $\mathcal{V}$*, and* $\mathcal{W}$ *are present in the antisymmetric inner product, as described in the following sections.*

IV.2 Antisymmetric Inner Product with $\delta(\gamma-\gamma_1)$ and ($\mathcal{T}$ and/or $\mathcal{V}$) Present

To compute antisymmetric inner products involving operators, we will modify formulas from Section III. The first (trivial) modification is to denote the variable of integration in those formulas by γ', so as not to confuse it with the variable γ in $\delta(\gamma-\gamma_1)$. Next we replace $|\mathbb{L}|$ with $|\mathbb{L}(\gamma)|$ given by (58). Using (57), we can express

$$\begin{aligned}\mathbb{L}(\gamma)^{-1} = (\mathbb{E}+\mathbf{e}_1(\mathbf{\Phi}(\gamma)-\mathbf{d})^*)^{-1} &= (\mathbb{E}\,(\mathbb{I}+\mathbf{d}(\mathbf{\Phi}(\gamma)-\mathbf{d})^*))^{-1} \\ &= (\mathbb{I}+\mathbf{d}(\mathbf{\Phi}(\gamma)-\mathbf{d})^*)^{-1}\,\mathbb{E}^{-1}\,. \end{aligned}\tag{61}$$

Using the Sherman-Morrisson Formula (see e.g. [21] and (B5) in Appendix B) we then have

$$\mathbb{L}(\gamma)^{-1} = \left(\mathbb{I}-\frac{\mathbf{d}(\mathbf{\Phi}(\gamma)-\mathbf{d})^*}{1+(\mathbf{\Phi}(\gamma)-\mathbf{d})^*\mathbf{d}}\right)\mathbb{E}^{-1} = \left(\mathbb{I}+\mathbf{d}\frac{(\mathbf{d}-\mathbf{\Phi}(\gamma))^*}{\mathbf{\Phi}(\gamma)^*\mathbf{d}}\right)\mathbb{E}^{-1}\,. \tag{62}$$

The vector of functions $\mathbf{\Theta}$, which was defined by $\mathbb{L}^{-1}\tilde{\mathbf{\Phi}}$, now depends on the variable γ in $\delta(\gamma-\gamma_1)$ as well as its own internal variable γ'. Replacing $\mathbb{L}^{-1}$ with (62) and $\tilde{\mathbf{\Phi}}$ with $\tilde{\mathbf{\Phi}}(\gamma')+\mathbf{e}_1(\delta(\gamma-\gamma')-\tilde{\phi}_1(\gamma'))$, we obtain

$$\mathbf{\Theta}(\gamma,\gamma') = \left(\mathbb{I}+\mathbf{d}\frac{(\mathbf{d}-\mathbf{\Phi}(\gamma))^*}{\mathbf{\Phi}(\gamma)^*\mathbf{d}}\right)\mathbb{E}^{-1}\left(\tilde{\mathbf{\Phi}}(\gamma')+\mathbf{e}_1(\delta(\gamma-\gamma')-\tilde{\phi}_1(\gamma'))\right)\,. \tag{63}$$

To compute it, we first compute the base case $\tilde{\mathbf{\Theta}}(\gamma') = \mathbb{E}^{-1}\tilde{\mathbf{\Phi}}(\gamma')$. Multiplying out (63) and noting $\mathbf{d}^*\tilde{\mathbf{\Theta}} = \mathbf{d}^*\mathbb{E}^{\ddagger}\tilde{\mathbf{\Phi}} = \tilde{\phi}_1$, we obtain

$$\begin{aligned}\mathbf{\Theta}(\gamma,\gamma') = \tilde{\mathbf{\Theta}}(\gamma') + \mathbf{d}\frac{\mathbf{d}^*\tilde{\mathbf{\Theta}}(\gamma') - \mathbf{\Phi}(\gamma)^*\tilde{\mathbf{\Theta}}(\gamma') + \delta(\gamma-\gamma') - \tilde{\phi}_1(\gamma')}{\mathbf{\Phi}(\gamma)^*\mathbf{d}} \\ = \tilde{\mathbf{\Theta}}(\gamma') - \mathbf{d}\frac{\mathbf{\Phi}(\gamma)^*\tilde{\mathbf{\Theta}}(\gamma') - \delta(\gamma-\gamma')}{\mathbf{\Phi}(\gamma)^*\mathbf{d}}\,. \end{aligned}\tag{64}$$

We are now ready to state our main formulas.

Proposition 15 *When* $\mathbb{E}$ *is nonsingular,*

$$\left\langle \delta(\gamma-\gamma_1)\prod_{i=2}^{N}\tilde{\phi}_i(\gamma_i), (\mathcal{T}+\mathcal{V})\prod_{i=1}^{N}\phi_i(\gamma_i)\right\rangle_{\mathcal{A}} \tag{65}$$

is equal to

$$\frac{|\mathbb{E}|}{N!}\left[\mathbf{\Phi}(\gamma)^*\left(\mathbf{d}\int(\mathcal{T}_*+\mathcal{V}_*)\left[\mathbf{\Phi}\right]^*\tilde{\mathbf{\Theta}}d\gamma'-\int(\mathcal{T}_*+\mathcal{V}_*)\left[\mathbf{\Phi}^*\mathbf{d}\right]\tilde{\mathbf{\Theta}}d\gamma'\right)\right.$$
$$\left.+(\mathcal{T}_*+\mathcal{V}_*)\left[\mathbf{\Phi}^*\mathbf{d}\right](\gamma)\right], \quad (66)$$

which can be computed with total cost $\mathcal{O}(N^3+N^2M)$.

Proof: To compute (65), we start with $\frac{|\mathbb{L}|}{N!}\int(\mathcal{T}_*+\mathcal{V}_*)\left[\mathbf{\Phi}\right]^*\mathbf{\Theta}d\gamma'$ from (52) and substitute in (58) and (64) to obtain

$$\frac{|\mathbb{E}|\,\mathbf{\Phi}(\gamma)^*\mathbf{d}}{N!}\int(\mathcal{T}_*+\mathcal{V}_*)\left[\mathbf{\Phi}\right](\gamma')^*\left(\tilde{\mathbf{\Theta}}(\gamma')-\mathbf{d}\frac{\mathbf{\Phi}(\gamma)^*\tilde{\mathbf{\Theta}}(\gamma')-\delta(\gamma-\gamma')}{\mathbf{\Phi}(\gamma)^*\mathbf{d}}\right)d\gamma'. \quad (67)$$

Distributing out and rearranging, we have

$$\frac{|\mathbb{E}|}{N!}\int\mathbf{\Phi}(\gamma)^*\mathbf{d}(\mathcal{T}_*+\mathcal{V}_*)\left[\mathbf{\Phi}\right]^*(\gamma')\tilde{\mathbf{\Theta}}(\gamma')-(\mathcal{T}_*+\mathcal{V}_*)\left[\mathbf{\Phi}\right](\gamma')^*\mathbf{d}\mathbf{\Phi}(\gamma)^*\tilde{\mathbf{\Theta}}(\gamma')$$
$$+(\mathcal{T}_*+\mathcal{V}_*)\left[\mathbf{\Phi}\right](\gamma')^*\mathbf{d}\delta(\gamma-\gamma')d\gamma', \quad (68)$$

which yields (66). Although in (62) and (64) we divide by $\mathbf{\Phi}^*\mathbf{d}$, which could be zero, this denominator cancels in the final expression, so we can argue by continuity that the final expression is still valid. One can also prove this directly by determining the nullspace of $\mathbb{L}$ and then using (54). □

Remark 16 *It is the term with pointwise multiplication,* $(\mathcal{T}_*+\mathcal{V}_*)\left[\mathbf{\Phi}^*\mathbf{d}\right]$ *in (66), that allows adaptive refinement around the nuclei in the numerical algorithm.*

To obtain the formulas when $\mathbb{E}$ is singular, we follow the same logic as in Section III.6.1. Denote the singular vectors in the nullspace of $\mathbb{E}$ by $\{(\tilde{\mathbf{u}}_i,\tilde{\mathbf{v}}_i)\}$.

Proposition 17 *When* $\mathbb{E}$ *has rank deficiency greater than one, (65) is zero. When* $\mathbb{E}$ *has rank deficiency one, (65) is equal to*

$$\frac{1}{|\mathbb{E}^\ddagger|N!}\mathbf{\Phi}(\gamma)^*\left(\mathbf{d}\int(\mathcal{T}_*+\mathcal{V}_*)\left[\mathbf{\Phi}^*\tilde{\mathbf{v}}_1\right]\tilde{\mathbf{u}}_1^*\tilde{\mathbf{\Phi}}d\gamma'-\tilde{\mathbf{v}}_1\int(\mathcal{T}_*+\mathcal{V}_*)\left[\mathbf{\Phi}^*\mathbf{d}\right]\tilde{\mathbf{u}}_1^*\tilde{\mathbf{\Phi}}d\gamma'\right), \quad (69)$$

which can be computed with total cost $\mathcal{O}(N^3+N^2M)$.

IV.3 Antisymmetric Inner Product with $\delta(\gamma-\gamma_1)$ and $\mathcal{W}$ Present

Conceptually the derivation if $\mathcal{W}$ is present in the inner product is the same and we obtain the following propositions.

Proposition 18 *When* $\mathbb{E}$ *is nonsingular,*

$$\left\langle\delta(\gamma-\gamma_1)\prod_{i=2}^{N}\tilde{\phi}_i(\gamma_i),\mathcal{W}\prod_{i=1}^{N}\phi_i(\gamma_i)\right\rangle_{\mathcal{A}} \quad (70)$$

is equal to

$$\frac{1}{2}\frac{|\mathbb{E}|}{N!}\Big[2\left(\mathbf{\Phi}(\gamma)^*\mathbf{d}\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\tilde{\mathbf{\Theta}}\right](\gamma)-\mathbf{\Phi}(\gamma)^*\mathcal{W}_{\mathcal{P}}\left[\tilde{\mathbf{\Theta}}\mathbf{\Phi}^*\mathbf{d}\right](\gamma)\right)$$
$$+\mathbf{\Phi}(\gamma)^*\left(\mathbf{d}\int\mathbf{\Phi}^*\tilde{\mathbf{\Theta}}\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\tilde{\mathbf{\Theta}}\right]-\mathbf{\Phi}^*\mathcal{W}_{\mathcal{P}}\left[\tilde{\mathbf{\Theta}}\mathbf{\Phi}^*\right]\tilde{\mathbf{\Theta}}d\gamma'\right.$$
$$\left.-2\int\tilde{\mathbf{\Theta}}\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\tilde{\mathbf{\Theta}}\right]\mathbf{\Phi}^*\mathbf{d}-\tilde{\mathbf{\Theta}}\mathbf{\Phi}^*\mathcal{W}_{\mathcal{P}}\left[\tilde{\mathbf{\Theta}}\mathbf{\Phi}^*\mathbf{d}\right]d\gamma'\right)\Big]\,, \quad (71)$$

which can be computed with total cost $\mathcal{O}(N^3+N^2M\log M)$.

Proposition 19 *When* $\mathbb{E}$ *has rank deficiency one, (70) is equal to*

$$\frac{1}{|\mathbb{E}^\ddagger|N!}\Big[\left(\mathbf{\Phi}(\gamma)^*\mathbf{d}\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\tilde{\mathbf{v}}_1\tilde{\mathbf{u}}_1^*\tilde{\mathbf{\Phi}}\right](\gamma)-\mathbf{\Phi}(\gamma)^*\tilde{\mathbf{v}}_1\mathcal{W}_{\mathcal{P}}\left[\tilde{\mathbf{u}}_1^*\tilde{\mathbf{\Phi}}\mathbf{\Phi}^*\mathbf{d}\right](\gamma)\right)$$
$$+\mathbf{\Phi}(\gamma)^*\left(\mathbf{d}\int\mathbf{\Phi}^*\tilde{\mathbf{v}}_1\left(\tilde{\mathbf{u}}_1^*\tilde{\mathbf{\Phi}}\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\tilde{\mathbf{\Theta}}\right]-\mathcal{W}_{\mathcal{P}}\left[\tilde{\mathbf{u}}_1^*\tilde{\mathbf{\Phi}}\mathbf{\Phi}^*\right]\tilde{\mathbf{\Theta}}\right)d\gamma'\right.$$
$$+\int\tilde{\mathbf{\Theta}}\left(\mathbf{\Phi}^*\tilde{\mathbf{v}}_1\mathcal{W}_{\mathcal{P}}\left[\tilde{\mathbf{u}}_1^*\tilde{\mathbf{\Phi}}\mathbf{\Phi}^*\mathbf{d}\right]-\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\tilde{\mathbf{v}}_1\tilde{\mathbf{u}}_1^*\tilde{\mathbf{\Phi}}\right]\mathbf{\Phi}^*\mathbf{d}\right)d\gamma'$$
$$\left.-\tilde{\mathbf{v}}_1\int\mathbf{\Phi}^*\mathbf{d}\left(\tilde{\mathbf{u}}_1^*\tilde{\mathbf{\Phi}}\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\tilde{\mathbf{\Theta}}\right]-\mathcal{W}_{\mathcal{P}}\left[\tilde{\mathbf{u}}_1^*\tilde{\mathbf{\Phi}}\mathbf{\Phi}^*\right]\tilde{\mathbf{\Theta}}\right)d\gamma'\right)\Big]\,, \quad (72)$$

which can be computed with total cost $\mathcal{O}(N^3+N^2M+NM\log M)$.

Proposition 20 *When* $\mathbb{E}$ *has rank deficiency two, (70) is equal to*

$$\frac{1}{|\mathbb{E}^\ddagger|N!}\mathbf{\Phi}(\gamma)^*\Big[\mathbf{d}\int\mathbf{\Phi}^*\tilde{\mathbf{v}}_1\tilde{\mathbf{u}}_1^*\tilde{\mathbf{\Phi}}\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\tilde{\mathbf{v}}_2\tilde{\mathbf{u}}_2^*\tilde{\mathbf{\Phi}}\right]-\mathbf{\Phi}^*\tilde{\mathbf{v}}_2\mathcal{W}_{\mathcal{P}}\left[\tilde{\mathbf{u}}_2^*\tilde{\mathbf{\Phi}}\mathbf{\Phi}^*\tilde{\mathbf{v}}_1\right]\tilde{\mathbf{u}}_1^*\tilde{\mathbf{\Phi}}d\gamma$$
$$-\tilde{\mathbf{v}}_1\int\mathbf{\Phi}^*\tilde{\mathbf{v}}_2\tilde{\mathbf{u}}_2^*\tilde{\mathbf{\Phi}}\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\mathbf{d}\tilde{\mathbf{u}}_1^*\tilde{\mathbf{\Phi}}\right]-\mathbf{\Phi}^*\tilde{\mathbf{v}}_2\mathcal{W}_{\mathcal{P}}\left[\tilde{\mathbf{u}}_2^*\tilde{\mathbf{\Phi}}\mathbf{\Phi}^*\mathbf{d}\right]\tilde{\mathbf{u}}_1^*\tilde{\mathbf{\Phi}}d\gamma$$
$$-\tilde{\mathbf{v}}_2\int\mathbf{\Phi}^*\tilde{\mathbf{v}}_1\tilde{\mathbf{u}}_1^*\tilde{\mathbf{\Phi}}\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\mathbf{d}\tilde{\mathbf{u}}_2^*\tilde{\mathbf{\Phi}}\right]-\mathbf{\Phi}^*\tilde{\mathbf{v}}_1\mathcal{W}_{\mathcal{P}}\left[\tilde{\mathbf{u}}_1^*\tilde{\mathbf{\Phi}}\mathbf{\Phi}^*\mathbf{d}\right]\tilde{\mathbf{u}}_2^*\tilde{\mathbf{\Phi}}d\gamma\Big]\,, \quad (73)$$

which can be computed with total cost $\mathcal{O}(N^3+NM+M\log M)$.

V Details of the Green's Function Iteration

In this section we fill in the missing pieces in the Green's function iteration algorithm outlined in Section II.2. First we give a representation for the Green's function itself. Then we use the methods in the previous sections to construct the vector $\mathbf{b}$ in (18) and the matrix $\mathbb{A}$ in (17) to form the normal equations (15). Next we give the algorithm from Section II.2 in outline form as pseudocode. Finally we gather the computational cost of the whole method, and present some linear algebra techniques to reduce it.

V.1 Representing the Green's Function

In this section we construct a separated representation for the Green's function $\mathcal{G}_\mu$ in (7), following the ideas in [4, 5] (see also [22, 23]). We will use this representation in Section V.2 when constructing the right-hand-side of the normal equations.

We begin by constructing an approximation of $1/t$ with exponentials such that

$$\left| \frac{1}{t} - \sum_{p=1}^{L} w_p \exp(-\tau_p t) \right| < \epsilon \,, \tag{74}$$

on the interval $t \in [1, \infty)$, with w_p and τ_p positive. Expansions of $1/t$ into exponentials have been used in several applications and constructed by diverse techniques; see [8, 29, 59, 6, 9, 24] and the references therein. The interval $[1, \infty)$ is addressed specifically in [9], where it is shown that the error rate $\epsilon = \mathcal{O}(\exp(-c\sqrt{L}))$ can be achieved, which means we can achieve $L = \mathcal{O}((\ln \epsilon)^2)$.

Substituting $t = s/(-\mu)$ for $\mu < 0$ into (74) and dividing by $-\mu$, one has

$$\left| \frac{1}{s} - \sum_{p=1}^{L} \frac{w_p}{-\mu} \exp(-\frac{\tau_p}{-\mu} s) \right| < \frac{\epsilon}{-\mu} \,, \tag{75}$$

valid on the interval $s \in [-\mu, \infty)$. In Fourier coordinates, we can express

$$\mathcal{G}_\mu = \frac{1}{2\pi^2 \sum \xi_i^2 - \mu} \,, \tag{76}$$

from which we see that $\|\mathcal{G}_\mu\| = 1/(-\mu)$. Since the denominator is at least $-\mu > 0$, we can substitute into (75) and obtain

$$\left| \mathcal{G}_\mu - \sum_{p=1}^{L} \frac{w_p}{-\mu} e^{-\tau_p} \bigotimes_{i=1}^{N} \exp(-\frac{2\pi^2 \tau_p}{-\mu} \xi_i^2) \right| < \frac{\epsilon}{-\mu} = \epsilon \|\mathcal{G}_\mu\| \,. \tag{77}$$

Thus we obtain an approximation of $\mathcal{G}_\mu$ with relative error ϵ in norm using L terms, with L independent of N and μ. To construct $\mathcal{G}_\mu$ as an integral operator in spatial coordinates, we apply the inverse Fourier transform to obtain

$$\mathcal{G}_\mu \approx \sum_{p=1}^{L} \bigotimes_{i=1}^{N} \mathcal{F}^p_{\mathbf{r}_i} \,, \tag{78}$$

where the convolution operator $\mathcal{F}^p_{\mathbf{r}_i}$, which depends implicitly on μ, is defined by

$$\begin{aligned} \mathcal{F}^p_{\mathbf{r}_i} f(\gamma_1, \ldots, \gamma_N) = \left(\frac{w_p}{-\mu e^{\tau_p}} \right)^{1/N} \left(\frac{-\mu}{2\pi\tau_p} \right)^{3/2} \times \\ \int \exp\left(-\frac{-\mu}{2\tau_p} \|\mathbf{r}_i - \mathbf{r}'\|^2 \right) f(\gamma_1, \ldots, \gamma_{i-1}, (\mathbf{r}', \sigma_i), \gamma_{i+1}, \ldots, \gamma_N) d\mathbf{r}' \,. \end{aligned} \tag{79}$$

This construction has theoretical value, since it has proved the following theorem.

Theorem 21 *For any $\epsilon > 0$, $\mu < 0$, and N, the N-particle Green's function $\mathcal{G}_\mu$ has a separated representation with relative error in operator norm bounded by ϵ using $L = \mathcal{O}((\ln \epsilon)^2)$ terms, with L independent of μ and N.*

V.2 Constructing the Right-Hand-Side Vector b in (18)

In order to do a step in the iteration, we need to construct the right-hand-side $\mathbf{b}$ in the normal equations (15) in Section II.2.2. Since $\mathcal{A}$ is an orthogonal projection, $\mathcal{A}$ and $\mathcal{G}_\mu$ commute, and $\mathcal{G}_\mu$ is self-adjoint, the entry (18) is equal to

$$b(l)(\gamma) = -\tilde{s}_l \sum_{m}^{r} s_m \left\langle \mathcal{A}\, \mathcal{G}_\mu \delta(\gamma - \gamma_1) \prod_{i=2}^{N} \tilde{\phi}_i^l(\gamma_i), [\mathcal{V} + \mathcal{W}] \prod_{i=1}^{N} \phi_i^m(\gamma_i) \right\rangle \,. \tag{80}$$

Substituting (78) in for $\mathcal{G}_\mu$ and rearranging, we have

$$b(l)(\gamma) = -\tilde{s}_l \sum_m^r s_m \sum_{p=1}^L \left\langle \mathcal{A}\mathcal{F}^p_{\mathbf{r}_1}\delta(\gamma-\gamma_1)\prod_{i=2}^N \mathcal{F}^p_{\mathbf{r}_i}\tilde{\phi}^l_i(\gamma_i), [\mathcal{V}+\mathcal{W}]\prod_{i=1}^N \phi^m_i(\gamma_i)\right\rangle . \tag{81}$$

The computation is of the same form for each value of the indices l, m, and p, so we can consider a single term and suppress the indices.

To evaluate a single term $\langle \mathcal{A}\mathcal{F}_{\mathbf{r}_1}\delta(\gamma-\gamma_1)\prod_{i=2}\mathcal{F}_{\mathbf{r}_i}\tilde{\phi}_i(\gamma_i), [\mathcal{V}+\mathcal{W}]\prod_{i=1}\phi_i(\gamma_i)\rangle$ we use the formulas in Propositions 15–20 in Sections IV.2 and IV.3, with two modifications. The first modification is that $\tilde{\mathbf{\Phi}}$ is replaced with $\mathcal{F}\tilde{\mathbf{\Phi}}$ throughout. This replacement causes no structural change to the formulas; it just changes the inputs. The second modification is caused by the replacement of $\delta(\gamma-\gamma_1)$ by $\mathcal{F}_{\mathbf{r}_1}\delta(\gamma-\gamma_1)$. The first row of $\mathbb{L}(\gamma)$ in (55) becomes $\mathcal{F}\mathbf{\Phi}(\gamma)^*$, which makes $|\mathbb{L}(\gamma)| = |\mathbb{E}|\,\mathcal{F}\mathbf{\Phi}(\gamma)^*\mathbf{d}$. Similarly, (64) becomes

$$\mathbf{\Theta}(\gamma,\gamma') = \tilde{\mathbf{\Theta}}(\gamma') - \mathbf{d}\frac{\mathcal{F}\mathbf{\Phi}(\gamma)^*\tilde{\mathbf{\Theta}}(\gamma') - \mathcal{F}\delta(\gamma-\gamma')}{\mathcal{F}\mathbf{\Phi}(\gamma)^*\mathbf{d}} . \tag{82}$$

Tracking $\mathcal{F}$ through the formulas, we find that all we need to do is to modify the formulas in Sections IV.2 and IV.3 by applying $\mathcal{F}$ to the final result.

V.3 Constructing the Matrix $\mathbb{A}$ in (17)

In this section we construct the kernels in (17) for the normal equations (15), using the same ideas as in Section IV. We fix l and l' and define

$$\begin{aligned}
K(\gamma,\gamma') &= \frac{A(l,l')(\gamma,\gamma')}{\tilde{s}_l\tilde{s}_{l'}} && (83)\\
\mathbf{w}(\gamma') &= \left[\begin{array}{ccc} \tilde{\phi}^l_2(\gamma') & \dots & \tilde{\phi}^l_N(\gamma') \end{array}\right]^* && (84)\\
\mathbf{y}(\gamma) &= \left[\begin{array}{ccc} \tilde{\phi}^{l'}_2(\gamma) & \dots & \tilde{\phi}^{l'}_N(\gamma) \end{array}\right]^* && (85)\\
\mathbb{D} &= \left[\begin{array}{ccc} \langle\tilde{\phi}^l_2,\tilde{\phi}^{l'}_2\rangle & \cdots & \langle\tilde{\phi}^l_2,\tilde{\phi}^{l'}_N\rangle \\ \vdots & \ddots & \vdots \\ \langle\tilde{\phi}^l_N,\tilde{\phi}^{l'}_2\rangle & \cdots & \langle\tilde{\phi}^l_N,\tilde{\phi}^{l'}_N\rangle \end{array}\right] . && (86)
\end{aligned}$$

Using Löwdin's rules (37) we have

$$K(\gamma,\gamma') = \frac{|\mathbb{L}|}{N!} = \frac{1}{N!}\left|\begin{array}{cc} \delta(\gamma-\gamma') & \mathbf{y}^*(\gamma) \\ \mathbf{w}(\gamma') & \mathbb{D} \end{array}\right| . \tag{87}$$

Expressing $\mathbb{L}$ as a low-rank perturbation of $\left[\begin{array}{cc} 1 & 0 \\ 0 & \mathbb{D} \end{array}\right]$, we have

$$\begin{aligned}
K(\gamma,\gamma') &= \frac{1}{N!}\left|\left[\begin{array}{cc} 1 & 0 \\ 0 & \mathbb{D} \end{array}\right] + \left[\begin{array}{c} 1 \\ 0 \end{array}\right]\left[\begin{array}{cc} 0 & \mathbf{y}^*(\gamma) \end{array}\right] + \left[\begin{array}{c} \delta(\gamma-\gamma')-1 \\ \mathbf{w}(\gamma') \end{array}\right]\left[\begin{array}{cc} 1 & 0 \end{array}\right]\right| \\
&= \frac{1}{N!}\left|\begin{array}{cc} 1 & 0 \\ 0 & \mathbb{D} \end{array}\right|\left|\mathbb{I} + \left[\begin{array}{c} 1 \\ 0 \end{array}\right]\left[\begin{array}{cc} 0 & \mathbf{y}^*(\gamma) \end{array}\right] + \left[\begin{array}{c} \delta(\gamma-\gamma')-1 \\ \mathbb{D}^{-1}\mathbf{w}(\gamma') \end{array}\right]\left[\begin{array}{cc} 1 & 0 \end{array}\right]\right| \\
&= \frac{|\mathbb{D}|}{N!}\left|\begin{array}{cc} 1 & \mathbf{y}^*(\gamma)\mathbb{D}^{-1}\mathbf{w}(\gamma') \\ 1 & \delta(\gamma-\gamma') \end{array}\right| = \frac{|\mathbb{D}|}{N!}\left(\delta(\gamma-\gamma') - \mathbf{y}^*(\gamma)\mathbb{D}^{-1}\mathbf{w}(\gamma')\right) . \qquad (88)
\end{aligned}$$

If $\mathbb{D}$ is singular then we apply the same logic as in Section III.6.1. If $\mathbb{D}$ has rank-deficiency greater than one then $K(\gamma,\gamma')=0$. If it has rank-deficiency one then we have $K(\gamma,\gamma')=$

$$\frac{1}{|\mathbb{D}^\ddagger|N!}\left|\mathbb{I}+\begin{bmatrix}0\\-\mathbf{v}\end{bmatrix}\begin{bmatrix}0 & \mathbf{v}^*\end{bmatrix}+\begin{bmatrix}1\\0\end{bmatrix}\begin{bmatrix}0 & \mathbf{y}^*(\gamma)\end{bmatrix}+\begin{bmatrix}\delta(\gamma-\gamma')-1\\ \mathbb{D}^\ddagger\mathbf{w}(\gamma')\end{bmatrix}\begin{bmatrix}1 & 0\end{bmatrix}\right|$$
$$=\frac{1}{|\mathbb{D}^\ddagger|N!}\begin{vmatrix}0 & 0 & \mathbf{v}^*\mathbb{D}^\ddagger\mathbf{w}(\gamma')\\ -\mathbf{y}^*(\gamma)\mathbf{v} & 1 & \mathbf{y}^*(\gamma)\mathbb{D}^\ddagger\mathbf{w}(\gamma')\\ 0 & 1 & \delta(\gamma-\gamma')\end{vmatrix}=\frac{-(\mathbf{y}^*(\gamma)\mathbf{v})(\mathbf{v}^*\mathbb{D}^\ddagger\mathbf{w}(\gamma'))}{|\mathbb{D}^\ddagger|N!}$$
$$=\frac{-(\mathbf{y}^*(\gamma)\mathbf{v})(\mathbf{u}^*\mathbf{w}(\gamma'))}{|\mathbb{D}^\ddagger|N!}, \quad (89)$$

where $\mathbb{D}^\ddagger$ is the modified pseudo-inverse of Definition 4.

In the nonsingular case, we can construct $\mathbb{D}$ at cost $\mathcal{O}(N^2M)$ and compute $\mathbb{D}^{-1}$ at cost $\mathcal{O}(N^3)$. Applying this kernel costs $\mathcal{O}(NM)$ to integrate against a function in γ', $\mathcal{O}(N^2)$ to apply $\mathbb{D}^{-1}$, and then $\mathcal{O}(NM)$ to apply $\mathbf{y}^*$ to the result. In the singular case, we can compute $\mathbb{D}^\ddagger$ at cost $\mathcal{O}(N^3)$ and construct $\mathbf{y}^*\mathbf{v}$ and $\mathbf{u}^*\mathbf{w}$ at cost $\mathcal{O}(NM)$. Since the variables separate, applying this kernel costs $\mathcal{O}(M)$.

Remark 22 *In the case $r=1$, which corresponds to the Hartree-Fock formulation, $\mathbb{D}=\mathbb{I}$ and $K(\gamma,\gamma')$ is just the projector orthogonal to $\{\tilde{\phi}_i\}_{i=2}^N$.*

V.4 Pseudocode

In this section we give the algorithm in outline form as pseudocode. We do not indicate when objects can be recalled or updated from previous computations.

Loop through I Green's function iterations (9,10,13). For each of these:
  Construct $\mathcal{G}_\mu$ as in Section V.1, obtaining the operators $\mathcal{F}^p$ in (79).
  Loop through the N directions (electrons). For each of these:
    Compute $A(l,l')$ via (88) for all (l,l').
    Compute $b(l)(\gamma)$ in (81) by:
      Loop in the r values of l and for each:
        Sum over the L values of p and for each:
          Compute $\mathcal{F}^p\phi_i^l$ for all i.
          Sum over the r values of m and for each:
            Using $\mathcal{F}^p\tilde{\mathbf{\Phi}}$ in place of $\tilde{\mathbf{\Phi}}$, construct $\mathbb{E}$ in (56).
            Compute $|\mathbb{E}|$ and $\mathbb{E}^{-1}$.
            Construct $\tilde{\mathbf{\Theta}}=\mathbb{E}^{-1}\mathcal{F}\tilde{\mathbf{\Phi}}$.
            Construct $\mathbf{\Phi}^*\tilde{\mathbf{\Theta}}$, $\mathbf{\Phi}^*\mathbf{d}$, and $\tilde{\mathbf{\Theta}}\mathbf{\Phi}^*$.
            Compute $\mathcal{W}_{\mathcal{P}}\left[\mathbf{\Phi}^*\tilde{\mathbf{\Theta}}\right]$ and $\mathcal{W}_{\mathcal{P}}\left[\tilde{\mathbf{\Theta}}\mathbf{\Phi}^*\right]$.
            Compute (66) and (71) using these ingredients.
            Apply $\mathcal{F}^p$ to $((66)+(71))$.
    Apply conjugate gradient to solve the normal equations (15).
  Renormalize as in (10).
  Update μ via (13).

Remark 23 *We have presented the algorithm in serial form for clarity. The loop in l, sum in p, and sum in m can be trivially parallelized. Parallelizing the loop through the N electrons would represent a change in the algorithm, which we will develop elsewhere.*

V.5 Overall Computational Cost

The computational cost is dominated by the repeated construction and solution of the normal equations (15). For a fixed direction, the construction cost is dominated by (81), which has r^2L inner products. The most costly portion of the inner products is (71), which requires $\mathcal{O}(N^3 + N^2M \log M)$ operations, giving us the net construction cost

$$\mathcal{O}(r^2LN^2(N + M \log M)) . \tag{90}$$

The operation count to solve the normal equations (15) by applying the matrix of integral operators $\mathbb{A}$ S times is

$$\mathcal{O}(r^2SN(N + M)) . \tag{91}$$

As we loop through the directions, we may reuse several quantities, so the total cost of the construction is less than N times the cost for one direction. In fact, the construction cost for the entire loop through N directions is of the same order as the cost for one direction. The application cost is simply multiplied by N. In the sections below we show how to update the construction for direction $k = 2$ using what we already have for direction $k = 1$, and then determine the cost for one loop through the directions. We defer the development of the technical linear algebra rules on low-rank updates to Appendix B, and here only show how to apply them to our problem. Our final conclusion is the computational cost

$$\mathcal{O}(Ir^2N^2[L(N + M \log M) + S(N + M)]), \tag{92}$$

where I the number of Green's function iterations.

V.5.1 Reuse in Computing $\mathbb{A}$

Let $\mathbb{D}_1$ denote $\mathbb{D}$ in (86) for directions one, and $\mathbb{D}_2$ the version for direction two. We let $\hat{\phi}_1^l$ denote the updated version of $\tilde{\phi}_1^l$. To construct $\mathbb{D}_2$ requires only the first column and row of $\mathbb{D}_1$ to be updated, specifically

$$\mathbb{D}_2 = \mathbb{D}_1 + \mathbf{e}_1 \begin{bmatrix} 0 & (\langle \hat{\phi}_1^l, \tilde{\phi}_3^{l'} \rangle - \langle \tilde{\phi}_2^l, \tilde{\phi}_3^{l'} \rangle) & \cdots \end{bmatrix} + \begin{bmatrix} \langle \hat{\phi}_1^l, \hat{\phi}_1^{l'} \rangle - \langle \tilde{\phi}_2^l, \tilde{\phi}_2^{l'} \rangle \\ \vdots \end{bmatrix} \mathbf{e}_1^* . \tag{93}$$

Computing those inner products involving $\hat{\phi}_1^l$ and $\hat{\phi}_1^{l'}$ costs $\mathcal{O}(NM)$. Using Proposition 24 twice, we compute $\mathbb{D}_2^\ddagger$, $|\mathbb{D}_2^\ddagger|$, and if appropriate $\mathbf{v}$, all at cost $\mathcal{O}(N^2)$. The formulas (87) and following are modified by inserting the extra column and row in the second place instead of the first, but otherwise the procedure is unchanged. The cost for one loop through the N directions is thus $\mathcal{O}(N^3 + N^2M)$.

V.5.2 Reuse in Computing Antisymmetric Inner Products with $\delta(\gamma - \gamma_1)$ and Operators

We again let $\hat{\phi}_1^l$ denote the updated version of $\tilde{\phi}_1^l$ computed during the $k = 1$ solve. The inner products needed to construct $\mathbb{E}_2$ require only the one row involving $\hat{\phi}_1$ to be updated, at cost $\mathcal{O}(NM)$. The vector $\mathbf{d}_1$ can be constructed by doing the SVD of $\mathbb{E}_1$ with the first row set to zero and then selecting one of the right singular vectors $\mathbf{v}_i$ with zero singular value. Using Proposition 24 we obtain the SVD of $\mathbb{E}_2$ with first row set to zero and second row containing the new inner products, and thus can find $\mathbf{d}_2$. Putting the first and second rows back in proper position, we then have

$$\begin{aligned} \mathbb{E}_2 = \mathbb{E}_1 + \mathbf{e}_1 \left(\begin{bmatrix} \langle \mathcal{F}\hat{\phi}_1, \phi_1 \rangle & \cdots & \langle \mathcal{F}\hat{\phi}_1, \phi_N \rangle \end{bmatrix} - \mathbf{d}_1^* \right) \\ + \mathbf{e}_2 \left(\mathbf{d}_2^* - \begin{bmatrix} \langle \mathcal{F}\tilde{\phi}_2, \phi_1 \rangle & \cdots & \langle \mathcal{F}\tilde{\phi}_2, \phi_N \rangle \end{bmatrix} \right) , \end{aligned} \tag{94}$$

and we can compute $|\mathbb{E}_2^\ddagger|$ and $\mathbb{E}_2^\ddagger$ using Proposition 24 twice, at cost $\mathcal{O}(N^2)$.

Proposition 24 produces a rank two update and we must apply it twice. For notational ease we will show how to use a rank one update applied once; the method easily extends. Assuming $\mathbb{E}_2^\ddagger = \mathbb{E}_1^\ddagger + \mathbf{f}\mathbf{g}^*$, we next update

$$\begin{aligned}\tilde{\mathbf{\Theta}}_2 = \mathbb{E}_2^\ddagger \mathcal{F}\tilde{\mathbf{\Phi}}_2 = (\mathbb{E}_1^\ddagger + \mathbf{f}\mathbf{g}^*)(\mathcal{F}\tilde{\mathbf{\Phi}}_1 + \mathbf{e}_1(\hat{\phi}_1 - \tilde{\phi}_1)) \\ = \tilde{\mathbf{\Theta}}_1 + \mathbf{d}_1(\hat{\phi}_1 - \tilde{\phi}_1) + \mathbf{f}\mathbf{g}^*\mathcal{F}\tilde{\mathbf{\Phi}}_1 + \mathbf{f}\mathbf{g}^*\mathbf{e}_1(\hat{\phi}_1 - \tilde{\phi}_1) \quad (95)\end{aligned}$$

at cost $\mathcal{O}(NM)$. It is insufficient to just update $\tilde{\mathbf{\Theta}}_2$ in this way, since it would still cost $\mathcal{O}(N^2 M \log M)$ to compute $\mathcal{W}_\mathcal{P}\left[\tilde{\mathbf{\Theta}}_2\mathbf{\Phi}^*\right]$ in (71). Instead we update the combined quantity

$$\begin{aligned}\mathbf{\Phi}^*\mathcal{W}_\mathcal{P}\left[\tilde{\mathbf{\Theta}}_2\mathbf{\Phi}^*\right] = \mathbf{\Phi}^*\mathcal{W}_\mathcal{P}\left[\tilde{\mathbf{\Theta}}_1\mathbf{\Phi}^*\right] + \mathbf{\Phi}^*\mathbf{d}_1\mathcal{W}_\mathcal{P}\left[(\hat{\phi}_1 - \tilde{\phi}_1)\mathbf{\Phi}^*\right] \\ + \mathbf{\Phi}^*\mathbf{f}\mathcal{W}_\mathcal{P}\left[\mathbf{g}^*\mathcal{F}\tilde{\mathbf{\Phi}}_1\mathbf{\Phi}^*\right] + \mathbf{\Phi}^*\mathbf{f}\mathbf{g}^*\mathbf{e}_1\mathcal{W}_\mathcal{P}\left[(\hat{\phi}_1 - \tilde{\phi}_1)\mathbf{\Phi}\right] \quad (96)\end{aligned}$$

at cost $\mathcal{O}(NM \log M)$. With this quantity and $\tilde{\mathbf{\Theta}}_2$ we can compute (71) at cost $\mathcal{O}(NM \log M)$. The singular cases work similarly. The cost for one loop through the N directions is thus $\mathcal{O}(N^2 M \log M)$.

Acknowledgments

We would like to thank Dr. Robert Harrison (U. of Tennessee and Oak Ridge National Lab) and Dr. Lucas Monzón (U. of Colorado) for many useful conversations. Portions of this research were conducted while G.B. and M.J.M. were in residence at the Institute for Pure and Applied Mathematics during the fall of 2004.

This material is based upon work supported by the National Science Foundation under Grants DMS-0219326 (G.B and M.J.M.), DMS-0612358 (G.B. and F.P.), and DMS-0545895 (M.J.M.), the DARPA/ARO under Grants W911NF-04-1-0281 and W911NF-06-1-0254 (G.B and M.J.M.), and the Department of Energy under Grants DE-FG02-03ER25583 and DOE/ORNL Grant 4000038129 (G.B. and F.P.).

A Appendix: Algorithms Based on Gradient Descent

We prefer the integral iteration in Section II.2.1 due to the generally superior numerical properties of integral formulations. One could, however, try to minimize (4) directly with a method based on gradients. Since the machinery that we have constructed applies to these methods as well, we sketch how it can be used.

To minimize (4) we could use a gradient descent, starting at some initial guess for ψ. Inserting our current approximation ψ and formally taking the gradient, we have

$$2\frac{\langle \mathcal{H}\psi, \nabla\psi\rangle_\mathcal{A}\langle\psi,\psi\rangle_\mathcal{A} - \langle\mathcal{H}\psi,\psi\rangle_\mathcal{A}\langle\psi,\nabla\psi\rangle_\mathcal{A}}{\langle\psi,\psi\rangle_\mathcal{A}^{\,2}}\,. \quad (A1)$$

Defining μ to be our current value of (4), the gradient reduces to

$$\frac{2}{\langle\psi,\psi\rangle_\mathcal{A}}\left(\langle\mathcal{H}\psi,\nabla\psi\rangle_\mathcal{A} - \mu\langle\psi,\nabla\psi\rangle_\mathcal{A}\right)\,. \quad (A2)$$

The gradient is with respect to the parameters that are used to minimize (4). In our case that is the values of the functions ϕ_j^l. Taking the gradient with respect to the point values

of ϕ_j^l results in a vector $\mathbf{g}$ of functions, defined by

$$g_j^l(\gamma) = \frac{2}{\langle \psi, \psi \rangle_{\mathcal{A}}} s_l \sum_{m=1}^{r} s_m \left\langle \delta(\gamma - \gamma_j) \prod_{i \neq j}^{N} \phi_i^l(\gamma_i), (\mathcal{H} - \mu \mathcal{I}) \prod_{i=1}^{N} \phi_i^m(\gamma_i) \right\rangle_{\mathcal{A}}, \tag{A3}$$

where $\delta(\gamma - \gamma_j)$ is the delta function. The methods in Section IV can be used to construct $\mathbf{g}$.

Moving t in the direction opposite the gradient replaces ψ with

$$\sum_{l=1}^{r} s_l \prod_{i=1}^{N} (\phi_i^l - t g_i^l). \tag{A4}$$

Some search procedure can then be used to find an appropriate t. Then ψ is updated and the procedure repeated.

Alternatively, we can use an alternating direction approach and take the gradient with respect to the functions ϕ_i^l for one direction i, while fixing the functions in the other directions, and then loop through the directions. This loop through the directions is then repeated I times until we obtain the desired accuracy. We describe the $i = 1$ case. Moving t in the direction opposite the gradient replaces ψ with

$$\sum_{l=1}^{r} s_l (\phi_1^l - t g_1^l) \prod_{i=2}^{N} \phi_i^l = \psi - t \sum_{l=1}^{r} s_l g_1^l \prod_{i=2}^{N} \phi_i^l = \psi - t\tilde{\psi}. \tag{A5}$$

Inserting (A5) into (4) results in

$$\frac{\left\langle \mathcal{H}(\psi - t\tilde{\psi}), \psi - t\tilde{\psi} \right\rangle_{\mathcal{A}}}{\left\langle \psi - t\tilde{\psi}, \psi - t\tilde{\psi} \right\rangle_{\mathcal{A}}} = \frac{\langle \mathcal{H}\psi, \psi \rangle_{\mathcal{A}} - 2t \left\langle \mathcal{H}\psi, \tilde{\psi} \right\rangle_{\mathcal{A}} + t^2 \left\langle \mathcal{H}\tilde{\psi}, \tilde{\psi} \right\rangle_{\mathcal{A}}}{\langle \psi, \psi \rangle_{\mathcal{A}} - 2t \left\langle \psi, \tilde{\psi} \right\rangle_{\mathcal{A}} + t^2 \left\langle \tilde{\psi}, \tilde{\psi} \right\rangle_{\mathcal{A}}}. \tag{A6}$$

Once the inner products have been computed, we can find the minimizer for (A6) by solving a quadratic equation, and then update ψ via (A5). The cost to construct $\mathbf{g}$ for one direction is r^2 times the cost for one inner product. The dominant cost for the inner product comes from (71), which costs $\mathcal{O}(N^3 + N^2 M \log M)$, giving us the net construction cost

$$\mathcal{O}(r^2 N^2 (N + M \log M)). \tag{A7}$$

As described in Section V.5.2, many of the computations can be reused, so the cost for a single loop through the N directions is of the same order. Thus, for I loops through the directions the overall computational cost is

$$\mathcal{O}(I r^2 N^2 (N + M \log M)). \tag{A8}$$

B Appendix: Low-rank Updates

In this section we develop formulas for low-rank updates to $\mathbb{A}^\dagger$, $\mathbb{A}^\perp$ and $|\mathbb{A}^\ddagger|$, based on [42, 3].

Proposition 24 *Given* $\mathbb{A}$, $\mathbb{A}^\dagger$, $\mathbb{A}^\perp$, $|\mathbb{A}^\ddagger|$, $\mathbf{b}$, *and* $\mathbf{c}$, *let* $\mathbb{A}_1 = \mathbb{A} + \mathbf{b}\mathbf{c}^*$ *and compute*

$$\begin{array}{cccc} \mathbf{d} = \mathbb{A}^\dagger \mathbf{b}, & \mathbf{e} = (\mathbb{A}^\dagger)^* \mathbf{c}, & \mathbf{f} = (\mathbb{I} - \mathbb{A}\mathbb{A}^\dagger)\mathbf{b}, & \mathbf{g} = (\mathbb{I} - \mathbb{A}^\dagger \mathbb{A})\mathbf{c}, \\ d = \mathbf{d}^* \mathbf{d}, & e = \mathbf{e}^* \mathbf{e}, & f = \mathbf{f}^* \mathbf{f}, & g = \mathbf{g}^* \mathbf{g}, \\ \lambda = 1 + \mathbf{c}^* \mathbb{A}^\dagger \mathbf{b}, & \mu = |\lambda|^2 + dg, & \nu = |\lambda|^2 + ef, & \\ \mathbf{p} = \bar{\lambda}\mathbf{d} + d\mathbf{g}, & \mathbf{q} = \lambda \mathbf{e} + e\mathbf{f}. & & \end{array} \tag{B1}$$

1. If $\lambda = 0$, $f = 0$, and $g = 0$, then $rank(\mathbb{A}_1) = rank(\mathbb{A}) - 1$ and

$$\mathbb{A}_1^\dagger = \mathbb{A}^\dagger - d^{-1}\mathbf{d}\mathbf{d}^*\mathbb{A}^\dagger + e^{-1}(-\mathbb{A}^\dagger\mathbf{e} + d^{-1}(\mathbf{d}^*\mathbb{A}^\dagger\mathbf{e})\mathbf{d})\mathbf{e}^* \tag{B2}$$

$$\mathbb{A}_1^\perp = \mathbb{A}^\perp + (1/\sqrt{de})\mathbf{d}\mathbf{e}^* \tag{B3}$$

$$|\mathbb{A}_1^\ddagger| = -(1/\sqrt{de})|\mathbb{A}^\ddagger|\,. \tag{B4}$$

2. If $\lambda \neq 0$, $f = 0$, and $g = 0$, then $rank(\mathbb{A}_1) = rank(\mathbb{A})$ and

$$\mathbb{A}_1^\dagger = \mathbb{A}^\dagger - \lambda^{-1}\mathbf{d}\mathbf{e}^* \tag{B5}$$

$$\mathbb{A}_1^\perp = \mathbb{A}^\perp \tag{B6}$$

$$|\mathbb{A}_1^\ddagger| = |\mathbb{A}^\ddagger|\lambda^{-1}\,. \tag{B7}$$

3. If $f = 0$ and $g \neq 0$, then $rank(\mathbb{A}_1) = rank(\mathbb{A})$ and

$$\mathbb{A}_1^\dagger = \mathbb{A}^\dagger - \mu^{-1}\mathbf{d}(g\mathbf{d}^*\mathbb{A}^\dagger + \bar{\lambda}\mathbf{e}^*) + \mu^{-1}\mathbf{g}(-d\mathbf{e}^* + \lambda\mathbf{d}^*\mathbb{A}^\dagger) \tag{B8}$$

$$\mathbb{A}_1^\perp = \mathbb{A}^\perp - \frac{|\lambda|(\sqrt{\mu} - |\lambda|)\mathbf{g} + \lambda g\mathbf{d}}{g|\lambda|\sqrt{\mu}}\mathbf{g}^*\mathbb{A}^\perp \tag{B9}$$

$$|\mathbb{A}_1^\ddagger| = |\mathbb{A}^\ddagger|\frac{(\bar{\lambda} - \lambda)|\lambda|^2 + \lambda\mu}{\mu|\lambda|\sqrt{\mu}}\,. \tag{B10}$$

4. If $f \neq 0$ and $g = 0$, then $rank(\mathbb{A}_1) = rank(\mathbb{A})$ and

$$\mathbb{A}_1^\dagger = \mathbb{A}^\dagger - \nu^{-1}(f\mathbb{A}^\dagger\mathbf{e} + \bar{\lambda}\mathbf{d})\mathbf{e}^* + \nu^{-1}(-e\mathbf{d} + \lambda\mathbb{A}^\dagger\mathbf{e})\mathbf{f}^* \tag{B11}$$

$$\mathbb{A}_1^\perp = \mathbb{A}^\perp - \mathbb{A}^\perp\mathbf{f}\frac{(|\lambda|(\sqrt{\nu} - |\lambda|)\mathbf{f} + \bar{\lambda}f\mathbf{e})^*}{f|\lambda|\sqrt{\nu}} \tag{B12}$$

$$|\mathbb{A}_1^\ddagger| = |\mathbb{A}^\ddagger|\frac{(\lambda - \bar{\lambda})|\lambda|^2 + \bar{\lambda}\nu}{\nu|\lambda|\sqrt{\nu}}\,. \tag{B13}$$

5. If $f \neq 0$ and $g \neq 0$, then $rank(\mathbb{A}_1) = rank(\mathbb{A}) + 1$ and

$$\mathbb{A}_1^\dagger = \mathbb{A}^\dagger - f^{-1}\mathbf{d}\mathbf{f}^* + g^{-1}\mathbf{g}(-\mathbf{e}^* + \lambda f^{-1}\mathbf{f}^*) \tag{B14}$$

$$\mathbb{A}_1^\perp = \mathbb{A}^\perp - (1/\sqrt{gf})\mathbf{g}\mathbf{f}^* \tag{B15}$$

$$|\mathbb{A}_1^\ddagger| = |\mathbb{A}^\ddagger|\left[1 + (g^{-1}f^{-1} - (1/\sqrt{gf}))\mathbf{g}^*\mathbb{A}^\perp\mathbf{f}\right]\,. \tag{B16}$$

The cost to compute $\mathbb{A}_1^\dagger$, $\mathbb{A}_1^\perp$, and $|\mathbb{A}_1^\ddagger|$ is $\mathcal{O}(N^2)$.

Proof: The overall method, update rules for $rank(\mathbb{A}_1)$, and update rules for $\mathbb{A}_1^\dagger$ are taken from [3], who also list the useful properties

$$\begin{gathered}\mathbf{c}^*\mathbf{d} = \mathbf{e}^*\mathbf{b} = \lambda - 1, \quad \mathbf{b}^*\mathbf{f} = f, \quad \mathbf{c}^*\mathbf{g} = g, \quad \mathbf{d}^*\mathbf{g} = 0, \quad \mathbf{e}^*\mathbf{f} = 0, \\ \mathbb{A}^\dagger\mathbb{A}\mathbf{d} = \mathbf{d}, \quad \mathbb{A}\mathbb{A}^\dagger\mathbf{e} = \mathbf{e}, \quad \mathbb{A}^*\mathbf{f} = \mathbb{A}^\dagger\mathbf{f} = 0, \quad \mathbb{A}\mathbf{g} = (\mathbb{A}^\dagger)^*\mathbf{g} = 0.\end{gathered} \tag{B17}$$

They give update rules for the row and column spans of $\mathbb{A}_1$, which we translate into update rules for $\mathbb{A}^\perp$. The cases (B3), (B6), and (B15) follow directly. Corresponding to (B9), their update rule is that the row span of $\mathbb{A}^\perp$ should be extended (orthogonally) by $\mathbf{d}$ and then reduced by projecting orthogonal to $\mathbf{p}$. We translate this into a (Householder) reflection of the vector $\mathbf{g}$ into a vector in the span of $\mathbf{d}$ and $\mathbf{g}$ perpendicular to $\mathbf{p}$. Adjusting these

vectors to have equal norm and real inner product yields the reflection of the vector $\bar{\lambda}\sqrt{\mu}\,\mathbf{g}$ to $-|\lambda|(g\mathbf{d}-\bar{\lambda}\mathbf{g})$, resulting in

$$\left(\mathbb{I}-\frac{2(\bar{\lambda}\sqrt{\mu}\mathbf{g}+|\lambda|(g\mathbf{d}-\bar{\lambda}\mathbf{g}))(\bar{\lambda}\sqrt{\mu}\mathbf{g}+|\lambda|(g\mathbf{d}-\bar{\lambda}\mathbf{g}))^*}{\|(\bar{\lambda}\sqrt{\mu}\mathbf{g}+|\lambda|(g\mathbf{d}-\bar{\lambda}\mathbf{g}))\|^2}\right)\mathbb{A}^{\perp}\,, \tag{B18}$$

which simplifies to (B9). To obtain (B12) we use the same process, extending the column span by $\mathbf{e}$ and then projecting orthogonal to $\mathbf{q}$ by a reflection of $\lambda\sqrt{\nu}\,\mathbf{f}$ to $-|\lambda|(f\mathbf{e}-\lambda\mathbf{f})$.

To derive the update rules for $|\mathbb{A}_1^{\ddagger}|$, first add the update rules for $\mathbb{A}_1^{\dagger}$ and $\mathbb{A}_1^{\perp}$ and then take the determinant. On the right hand side factor out a copy of $\mathbb{A}^{\ddagger}$ leaving a low-rank perturbation of the identity, to which we can apply Proposition 2. To simplify the results, we use (B1), (B17), and the further observations

$$\begin{array}{ccc}(\mathbb{A}^{\ddagger})^{-1}\mathbf{d}=\mathbf{b}-\mathbf{f}, & (\mathbb{A}^{\ddagger})^{-1}\mathbb{A}^{\dagger}\mathbf{e}=\mathbf{c}-\mathbf{g}, & (\mathbb{A}^{\ddagger})^{-1}\mathbf{g}=(\mathbb{A}^{\perp})^*\mathbf{c},\\ \mathbf{e}^*(\mathbb{A}^{\ddagger})^{-1}=\mathbf{c}^*-\mathbf{g}^*, & \mathbf{f}^*(\mathbb{A}^{\ddagger})^{-1}=\mathbf{b}^*(\mathbb{A}^{\perp})^*. & \end{array} \tag{B19}$$

To obtain (B4) we compute

$$\begin{aligned}|\mathbb{A}_1^{\ddagger}| &= |\mathbb{A}^{\ddagger}|\left|\mathbb{I}-d^{-1}\mathbf{b}\mathbf{d}^*\mathbb{A}^{\dagger}+((1/\sqrt{de})\mathbf{b}-e^{-1}\mathbf{e}+d^{-1}e^{-1}(\mathbf{d}^*\mathbb{A}^{\dagger}\mathbf{e})\mathbf{b})\mathbf{e}^*\right|\\ &= |\mathbb{A}^{\ddagger}|\begin{vmatrix} 1-d^{-1}\mathbf{d}^*\mathbb{A}^{\dagger}\mathbf{b} & \mathbf{d}^*\mathbb{A}^{\dagger}((1/\sqrt{de})\mathbf{b}-e^{-1}\mathbf{e}+d^{-1}e^{-1}(\mathbf{d}^*\mathbb{A}^{\dagger}\mathbf{e})\mathbf{b})\\ -d^{-1}\mathbf{e}^*\mathbf{b} & 1+\mathbf{e}^*((1/\sqrt{de})\mathbf{b}-e^{-1}\mathbf{e}+d^{-1}e^{-1}(\mathbf{d}^*\mathbb{A}^{\dagger}\mathbf{e})\mathbf{b})\end{vmatrix}\\ &= |\mathbb{A}^{\ddagger}|\begin{vmatrix} 0 & \mathbf{d}^*\mathbb{A}^{\dagger}(1/\sqrt{de})\mathbf{b}\\ d^{-1} & \mathbf{e}^*((1/\sqrt{de})\mathbf{b}+d^{-1}e^{-1}(\mathbf{d}^*\mathbb{A}^{\dagger}\mathbf{e})\mathbf{b})\end{vmatrix} = |\mathbb{A}^{\ddagger}|(-(1/\sqrt{de}))\,.\end{aligned} \tag{B20}$$

For (B7) we have $|\mathbb{A}_1^{\ddagger}| = |\mathbb{A}^{\ddagger}|\left|\mathbb{I}-\lambda^{-1}\mathbf{b}\mathbf{e}^*\right| = |\mathbb{A}^{\ddagger}|(1-\lambda^{-1}\mathbf{e}^*\mathbf{b}) = |\mathbb{A}^{\ddagger}|\lambda^{-1}$. To obtain (B10) we compute

$$\begin{aligned}|\mathbb{A}^{\ddagger}|\Big|\mathbb{I}+(\mathbb{A}^{\ddagger})^{-1}\Big(&\mathbf{d}(-\mu^{-1}(g\mathbf{d}^*\mathbb{A}^{\dagger}+\bar{\lambda}\mathbf{e}^*)-\frac{\lambda\mathbf{g}^*\mathbb{A}^{\perp}}{|\lambda|\sqrt{\mu}})\\ &+\mathbf{g}(\mu^{-1}(-d\mathbf{e}^*+\lambda\mathbf{d}^*\mathbb{A}^{\dagger})-\frac{(\sqrt{\mu}-|\lambda|)\mathbf{g}^*\mathbb{A}^{\perp}}{g\sqrt{\mu}})\Big)\Big|\\ = |\mathbb{A}^{\ddagger}|&\begin{vmatrix} 1+(-\mu^{-1}(g\mathbf{d}^*\mathbb{A}^{\dagger}+\bar{\lambda}\mathbf{e}^*))\mathbf{b} & (\mu^{-1}(-d\mathbf{e}^*+\lambda\mathbf{d}^*\mathbb{A}^{\dagger}))^*\mathbf{b}\\ (-\lambda\mathbf{g}^*\mathbb{A}^{\perp}/|\lambda|\sqrt{\mu})(\mathbb{A}^{\perp})^*\mathbf{c} & 1-((\sqrt{\mu}-|\lambda|)\mathbf{g}^*\mathbb{A}^{\perp}/g\sqrt{\mu})^*(\mathbb{A}^{\perp})^*\mathbf{c}\end{vmatrix}\\ = |\mathbb{A}^{\ddagger}|&\begin{vmatrix}\bar{\lambda}/\mu & d/\mu\\ -\lambda g/|\lambda|\sqrt{\mu} & |\lambda|/\sqrt{\mu}\end{vmatrix} = |\mathbb{A}^{\ddagger}|\frac{(\bar{\lambda}-\lambda)|\lambda|^2+\lambda\mu}{\mu|\lambda|\sqrt{\mu}}\,.\end{aligned} \tag{B21}$$

A similar calculation yields (B13). To obtain (B16) we compute

$$\begin{aligned}|\mathbb{A}^{\ddagger}|\Big|\mathbb{I}+(\mathbb{A}^{\ddagger})^{-1}\Big(-f^{-1}\mathbf{d}\mathbf{f}^*+\mathbf{g}(g^{-1}(-\mathbf{e}^*+\lambda f^{-1}\mathbf{f}^*)-(1/\sqrt{gf})\mathbf{f}^*)\Big)\Big|&\\ = |\mathbb{A}^{\ddagger}|\begin{vmatrix} 1 & \mathbf{f}^*(\mathbb{A}^{\perp})^*\mathbf{c}\\ g^{-1}f^{-1}(\lambda-1) & 1+(g^{-1}\lambda f^{-1}-(1/\sqrt{gf}))\mathbf{f}^*(\mathbb{A}^{\perp})^*\mathbf{c}\end{vmatrix}&\\ = |\mathbb{A}^{\ddagger}|(1+((-(1/\sqrt{gf}))+g^{-1}f^{-1})\mathbf{f}^*(\mathbb{A}^{\perp})^*\mathbf{c})&\\ = |\mathbb{A}^{\ddagger}|\left[1+(g^{-1}f^{-1}-(1/\sqrt{gf}))\mathbf{g}^*\mathbb{A}^{\perp}\mathbf{f}\right].&\end{aligned} \tag{B22}$$

□

When $\mathbb{A}$ and $\mathbb{A}_1$ are nonsingular, (B5) is the Sherman-Morrisson Formula (see e.g. [21]). For our application we need the singular vectors in $\mathbb{A}^{\perp}$, rather than $\mathbb{A}^{\perp}$ itself, but then only when $rank(\mathbb{A}^{\perp}) \leq 3$. These singular vectors can be extracted by a simple modification of the power method with deflation.

References

[1] H. Agren, A. Flores-Riveros, and H.J.Aa. Jensen. Evaluation of first- and second-order nonadiabatic coupling elements from large multiconfigurational self-consistent-field wave functions. *Physical Review A*, 34(6):4606–4614, December 1986.

[2] Philippe Y. Ayala and H. Bernhard Schlegel. A nonorthogonal CI treatment of symmetry breaking in sigma formyloxyl radical. *J. Chem. Phys.*, 108(18):7560–7567, May 1998.

[3] Jerzy K. Baksalary, Oskar Maria Baksalary, and Gtz Trenkler. A revisitation of fomulae for the Moore-Penrose inverse of modified matrices. *Linear Algebra Appl.*, 372:207–224, 2003.

[4] G. Beylkin and M. J. Mohlenkamp. Numerical operator calculus in higher dimensions. *Proc. Natl. Acad. Sci. USA*, 99(16):10246–10251, August 2002. `http://www.pnas.org/cgi/content/abstract/112329799v1`.

[5] G. Beylkin and M. J. Mohlenkamp. Algorithms for numerical analysis in high dimensions. *SIAM J. Sci. Comput.*, 26(6):2133–2159, July 2005. `http://amath.colorado.edu/pub/wavelets/papers/BEY-MOH2005.pdf`.

[6] G. Beylkin and L. Monzón. On approximation of functions by exponential sums. *Appl. Comput. Harmon. Anal.*, 19(1):17–48, 2005. `http://amath.colorado.edu/pub/wavelets/papers/afes.pdf`.

[7] Gregory Beylkin, Vani Cheruvu, and Fernando Pérez. Fast adaptive algorithms in the non-standard form for multidimensional problems. *Appl. Comput. Harmon. Anal.*, accepted for publication, 2007. APPM Preprint #550, `http://arxiv.org/abs/0706.0747`.

[8] D. Braess. Asymptotics for the approximation of wave functions by exponential sums. *J. Approx. Theory*, 83(1):93–103, 1995.

[9] Dietrich Braess and Wolfgang Hackbusch. Approximation of $1/x$ by exponential sums in $[1, \infty)$. *IMA J. Numer. Anal.*, 25(4):685–697, 2005.

[10] Rasmus Bro. Parafac. tutorial & applications. In *Chemom. Intell. Lab. Syst., Special Issue 2nd Internet Conf. in Chemometrics (INCINC'96)*, volume 38, pages 149–171, 1997. `http://www.models.kvl.dk/users/rasmus/presentations/parafac_tutorial/paraf.htm`.

[11] Eric Cancès, Mireille Defranceschi, Werner Kutzelnigg, Claude Le Bris, and Yvon Maday. Computational quantum chemistry: a primer. In *Handbook of Numerical Analysis, Vol. X*, pages 3–270. North-Holland, Amsterdam, 2003.

[12] Bin Chen and James R. Anderson. A simplified released-node quantum monte carlo calculation of the ground state of LiH. *J. Chem. Phys.*, 102(11):4491–4494, March 1995.

[13] E.U. Condon and G.H. Shortley. *The Theory of Atomic Spectra*. Cambridge Uinversity press, 1967.

[14] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. On the best rank-1 and rank-$(R_1, R_2, \cdots, R_N)$ approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.*, 21(4):1324–1342, 2000.

[15] Fokke Dijkstra and Joop H. Van Lenthe. On the rapid evaluation of cofactors in the calculation of nonorthogonal matrix elements. *International Journal of Quantum Chemistry*, 67:77–83, 1998.

[16] Fokke Dijkstra and Joop H. Van Lenthe. Gradients in valence bond theory. *Journal of Chemical Physics*, 113(6):2100–2108, August 2000.

[17] Frank Ethridge and Leslie Greengard. A new fast-multipole accelerated Poisson solver in two dimensions. *SIAM J. Sci. Comput.*, 23(3):741–760 (electronic), 2001.

[18] Reinhold Fink and Volker Staemmler. A multi-configuration reference CEPA method based on pair natural orbitals. *Theoretica Chimica Acta*, 87(1/2):129–145, November 1993.

[19] P. Gilbert. The reconstruction of a three-dimensional structure from projections and its applications to electron microscopy II. Direct methods. *Proc. R. Soc. Lond. B.*, pages 89–102, 1972.

[20] T. L. Gilbert. Multiconfiguration self-consistent-field theory for localized orbitals. I. The orbital equations. *Phys. Rev. A*, 6(2), August 1972.

[21] G. Golub and C. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 3rd edition, 1996.

[22] W. Hackbusch and B. N. Khoromskij. Low-rank Kronecker-product approximation to multi-dimensional nonlocal operators. I. Separable approximation of multi-variate functions. *Computing*, 76(3-4):177–202, 2006.

[23] W. Hackbusch and B. N. Khoromskij. Low-rank Kronecker-product approximation to multi-dimensional nonlocal operators. II. HKT representation of certain operators. *Computing*, 76(3-4):203–225, 2006.

[24] Wolfgang Hackbusch. Entwicklungen nach exponentialsummen. Technical Report 4, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig, Germany, 2005. see also `http://www.mis.mpg.de/scicomp/EXP_SUM/`.

[25] R.J. Harrison, G.I. Fann, T. Yanai, and G. Beylkin. Multiresolution quantum chemistry in multiwavelet bases. In P.M.A. Sloot et. al., editor, *Lecture Notes in Computer Science. Computational Science-ICCS 2003*, volume 2660, pages 103–110. Springer, 2003.

[26] R.J. Harrison, G.I. Fann, T. Yanai, Z. Gan, and G. Beylkin. Multiresolution quantum chemistry: basic theory and initial applications. *J. Chem. Phys.*, 121(23):11587–11598, 2004. `http://amath.colorado.edu/pub/wavelets/papers/mrqc.pdf`.

[27] Richard A. Harshman. Foundations of the parafac procedure: Model and conditions for an "explanatory" multi-mode factor analysis. Working Papers in Phonetics 16, UCLA, 1970. `http://publish.uwo.ca/~harshman/wpppfac0.pdf`.

[28] T. Helgaker and P.R. Taylor. *Modern Electronic Structure Theory*. World Scientific, Singapore, 1995.

[29] Tomasz Hrycak and Vladimir Rokhlin. An improved fast multipole algorithm for potential fields. *SIAM J. Sci. Comput.*, 19(6):1804–1826 (electronic), 1998.

[30] Walter Hunziker. On the spectra of Schrödinger multiparticle Hamiltonians. *Helv. Phys. Acta*, 39:451–462, 1966.

[31] M. H. Kalos. Monte Carlo calculations of the ground state of three- and four-body nuclei. *Phys. Rev. (2)*, 128:1791–1795, 1962.

[32] M. H. Kalos. Monte Carlo integration of the Schrödinger equation. *Trans. New York Acad. Sci. (2)*, 26:497–504, 1963/1964.

[33] Tosio Kato. Fundamental properties of Hamiltonian operators of Schrödinger type. *Trans. Amer. Math. Soc.*, 70:195–211, 1951.

[34] W. Klopper. R12 methods, Gaussian geminals. In J. Grotendorst et. al., editor, *Modern Methods and Algorithms of Quantum Chemistry*, volume 1 of *NIC Series*, pages 153–201. John von Neumann Institute for Computing, 2000.

[35] Wim Klopper and Claire C. M. Samson. Explicitly correlated second-order Møller Plesset methods with auxiliary basis sets. *Journal of Chemical Physics*, 116(15), April 2002.

[36] Pieter M. Kroonenberg and Jan de Leeuw. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45(1):69–97, 1980.

[37] C. Le Bris, editor. *Handbook of Numerical Analysis. Vol. X*. North-Holland, Amsterdam, 2003. Special Volume: Computational Chemistry.

[38] S. E. Leurgans, R. A. Moyeed, and B. W. Silverman. Canonical correlation analysis when the data are curves. *J. Roy. Statist. Soc. Ser. B*, 55(3):725–740, 1993.

[39] Roland Lindh, Jeppe Olsen, and Björn O. Roos. Low-rank configuration interaction with orbital optimization—the LR SCF approach. *Chemical Physics Letters*, 148(4):276–280, July 1988.

[40] Per-Olov Löwdin. Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configuration interaction. *Physical Review*, 97(6):1474–1489, March 1955.

[41] Arne Lüchow and Reinhold Fink. On the systematic improvement of fixed-node diffusion quantum Monte Carlo energies using natural orbital CI guide functions. *J. Chem. Phys.*, 113(19):8457–8463, November 2000.

[42] Carl D. Meyer, Jr. Generalized inversion of modified matrices. *SIAM J. Appl. Math.*, 24:315–323, 1973.

[43] Martin J. Mohlenkamp and Lucas Monzón. Trigonometric identities and sums of separable functions. *The Mathematical Intelligencer*, 27(2):65–69, 2005. `http://www.math.ohiou.edu/~mjm/research/sine.pdf`.

[44] Martin J. Mohlenkamp and Todd Young. Convergence of Green iterations for Schrödinger equations. In Xiaoping Shen, editor, *Proceedings of International Workshop on Computational Science and its Education 2005*, (to appear).

[45] Thomas Muir. *A Treatise on the Theory of Determinants*. Privately published, Albany New York, 1930. revised and enlarged by William H. Metzler.

[46] Jozef Noga, Werner Kutzelnigg, and Wim Klopper. CC-R12, a correlation cusp corrected coupled-cluster method with a pilot application to the Be_2 potential curve. *Chemical Physics Letters*, 199(5), November 1992.

[47] Jeppe Olsen, Per-Ake Malmquist, Björn O. Roos, Roland Lindh, and Per-Olaf Widmark. A non-linear approach to configuration interaction. The low-rank CI method (LR CI). *Chemical Physics Letters*, 133(2):91–101, January 1987.

[48] Ruben Pauncz. *The Symmetric Group in Quantum Chemistry*. CRC Press, Boca Raton, FL, 1995.

[49] B. Joakim Persson and Peter R. Taylor. Accurate quantum-chemical calculations: The use of gaussian-type geminal functions in the treatment of electron correlation. *J. Chem. Phys.*, 105(14):5915–5926, October 1996.

[50] B. Joakim Persson and Peter R. Taylor. Molecular integrals over gaussian-type geminal basis functions. *Theor. Chem. Acc.*, 97:240–250, 1997.

[51] V.V. Prasolov. *Problems and Theorems in Linear Algebra*, volume 134 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, R.I., 1994.

[52] Michael Reed and Barry Simon. *Methods of Modern Mathematical Physics. II. Fourier analysis, self-adjointness*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1975.

[53] Michael Reed and Barry Simon. *Methods of Modern Mathematical Physics. IV. Analysis of operators*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1978.

[54] Franz Rellich. Störungstheorie der Spektralzerlegung. V. *Math. Ann.*, 118:462–484, 1942.

[55] Sven Peter Rudin. *Configuration Interaction with Non-orthogonal Slater Determinants Applied to the Hubbard Model, Atoms, and Small Molecules*. PhD thesis, The Ohio State University, 1997.

[56] Jacek Rychlewski, Wojciech Cencek, and Jacek Komasa. The equivalence of explicitly correlated slater and gaussian functions in variational quantum chemistry computations. the ground state of H_2. *Chemical Physics Letters*, 229:657–660, November 1994.

[57] C. David Sherrill and Henry F. Schaefer III. The configuration interaction method: Advances in highly correlated approaches. *Advances in Quantum Chemisty*, 127:143–269, 1999.

[58] Age Smilde, Rasmus Bro, and Paul Geladi. *Multi-way Analysis. Applications in the Chemical Sciences*. John Wiley & Sons, 2004.

[59] N. Yarvin and V. Rokhlin. Generalized Gaussian quadratures and singular value decompositions of integral operators. *SIAM J. Sci. Comput.*, 20(2):699–718 (electronic), 1999.

[60] Jürgen Zanghellini. *Multi-Electron Dynamics in the Ionization of Molecules by Strong Laser Pulses*. PhD thesis, Vienna University of Technology, Vienna, Austria, 2004.